Computational Design and Experimental Validation of Selective Inhibitors of Enzymes Regulating Post-Translational Modifications

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Emilie Frugier

aus

Frankreich

Promotionskomitee

Prof. Dr. Amedeo Caflisch (Vorsitz)

Prof. Dr. Cristina Nevado

Prof. Dr. Michael Hottiger

Zürich 2014

ERKLÄRUNG

Die Dissertation wurde selbständig, ohne unerlaubte Hilfe angefertigt. Bei der Abfassung der Dissertation wurde im Sinne von §3 lit. 1 der Promotionsordnung vom 8. Juli 2002 keine anderen als die angegebenen Hilfsmittel verwendet.

Zürich, im Juni 2014

Emilie Frugier

ACKNOWLEDGEMENTS

I would like to express my genuine gratitude to all who accompanied and supported me during this thesis. The first person I would like to thank is my supervisor, Prof. Dr. Amedeo Caflisch, for giving me the opportunity to conduct my PhD thesis in his group, and for his encouragement, guidance and continuous support. I also want to sincerely thank the members of this thesis committee: Prof. Dr. Cristina Nevado of the Department of Chemistry and Prof. Dr. Michael Hottiger of the Institute of Veterinary Biochemistry and Molecular Biology, both at the University of Zurich.

As part of this PhD thesis was carried out in collaboration with others, I owe my further thanks to Dr. Florian Rosenthal and Dr. Paul Hassa, as well as Andrea Unzue and Yuri Pevzner.

Furthermore, I am thankful to all of the members of the Caflisch research group, and especially the “dockers” for many helpful discussions and Dr. Danzhi Huang for re-reading this introduction. My appreciation is further directed to the administration, and especially Christiane Gujan and Serena Caflisch, as well Dr. Stephan Klauser and Stever Rast of IT Services at the Department of Biochemistry. Thank you also to Ines Ehrnstorfer and Yvonne Neldner for their friendship and support.

My greatest thanks must be extended to my partner Eivind Helland, and my children, Mathias and Eliana, who supported my presence and absence with patience throughout this time. I would also like to thank my parents, Judith and Bernard Frugier, without whose support this would not have been possible.

This thesis is dedicated to my beloved aunt Sylvia Ruth Yonke, who died of metastatic breast cancer during my last PhD year. Our efforts to uncover new drug targets and the molecular entities modulating their actions will never have been fast enough.

CONTENTS

Erklärung i
Acknowledgements ii
Contents iii
Research Summary 1
Zusammenfassung 2

1. Introduction 3
- 1.1. Drug development process 3
  - 1.1.1 Target identification 4
  - 1.1.2 Small-molecule hit compound identification 5
  - 1.1.3 Small-molecule hit-to-lead derivatisation 9
  - 1.1.4 Small-molecule compound *in vivo* evaluation 9
- 1.2 Application of *in silico* methods to preclinical drug development 12
  - 1.2.1 Docking 13
  - 1.2.2 Scoring 15
  - 1.2.3 Digital Libraries 18
  - 1.2.4 Homology modelling 19
  - 1.2.5 Molecular dynamics 20
- 1.3 Aims and structure of this thesis 24
- 1.4 References 26

2. *In silico* identification of ARTD-1 selective inhibitors 29
- 2.1 Introduction 29
  - 2.1.1 ARTD family of proteins 29
  - 2.1.2 ARTD-1 and -2: similar proteins, differing roles? 31
  - 2.1.3 Overview of known ARTD-1 and ARTD-2 inhibitors 31
  - 2.1.4 Project objectives 31
- 2.2 ARTD-1/-2 binding site comparison 32
  - 2.2.1 Sequence comparison and structural alignment 32
  - 2.2.2 Residue accessibility to binding site: a molecular dynamics investigation 33
- 2.3 ARTD-1 protein conformations for docking 37
  - 2.3.1 Cut-based free energy profile 37
  - 2.3.2 Binding site residue dihedral angles and SASA within basins 37
  - 2.3.3 Summary of selected protein conformations 38
- 2.4 ARTD-1 scoring function 39
  - 2.4.1 Training set ligands and pose generation 39
  - 2.4.2 Evaluation of terms 40
  - 2.4.3 Scoring function: correlation and cross-validation correlation 41
  - 2.4.4 Score cut-off determination 42
- 2.5 Fragment-based library docking 42
  - 2.5.1 Motivation and protocol 42
  - 2.5.2 Fragment library generation 42
  - 2.5.3 Fragment docking, ranking and selection 43
  - 2.5.4 Fragment-containing ligand library docking and selection 44
  - 2.5.5 Summary of selected ligands 44
- 2.6 Nicotinamide-mimic library docking 45
  - 2.6.1 Motivation and protocol 45
  - 2.6.2 Nicotinamide-mimic library generation 46
  - 2.6.3 Ligand docking and selection 46
- 2.7 Discussion 48
- 2.8 Conclusions 48
- 2.8 References 49

3. Repurposing approved drugs and safe inhibitor 52
3.1 Introduction 52
  3.1.1 Repurposing motivation 52
  3.1.2 Project objectives 52
3.2 Compound libraries 52
  3.2.1 FDA-approved compounds 52
  3.2.2 Post-clinical trials phase I kinase inhibitors 53
3.3. Docking to ARTD-1 53
  3.3.1 Protocol 53
  3.3.2 Results and selected compounds 53
3.4 Conclusions 54
3.5 References 55

4. Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases 56
Rosenthal, F., Feijs, K., Frugier, E., Bonalli, M., Forst, A.H., Imhof, R., Winkler, H.C., Fischer, D., Caflish, A., Hassa, P.O., Lüscher, B. and Hottiger, M. *Nat. Struct. Mol. Biol.*, **2013**, 20(4), 502-509

5. Pyrrolo[3,2-b]quinoxaline-derivatives as Type I1/2 and II Eph Tyrosine Kinase Inhibitors: Structure-based Design, Synthesis, and In-vivo Validation 78
Unzue, A., Dong, J., Lafleur, K., Zhao, H., Frugier, E., Caflisch, A. and Nevado C. *J. Med. Chem.,* **2014**, 57, 6834-6844

Appendix A: Fragment-based docking: Development of the CHARMMing Web user Interface as a Platform for Computer-Aided Drug Design 90
Pevzner, Y., Frugier, E., Schalk, V., Caflisch, A. and Woodcock, H.L., *J. Chem. Inf. Model.,* **2014**, 54(9), 2612-2620

Appendix B: Curriculum vitae 100

RESEARCH SUMMARY

In silico methods can be applied in many ways to partly replace, complement or render more efficient experimental approaches to drug design. In this thesis, computational approaches were employed to:

- Determine physico-chemical differences between the binding sites of two related proteins, human ARTD-1 and ARTD-2, taking into account the intrinsic flexibility of these enzymes as observed by molecular dynamics simulations;
- Based on molecular dynamics simulations, identify protein conformations falling within densely explored regions of phase space ("energetically-favourable" basins of kinetically clustered conformations) of ARTD-1 that present physico-chemical differences relative to ARTD-2 for use in a high-throughput *in silico* docking campaign;
- Define a scoring function to apply to small-molecule compound poses within the ARTD-1 binding site with the aim of enriching the fraction of high binding affinity compounds of a library above a selected score cut-off;
- Propose for experimental validation small-molecule compounds binding ARTD-1 with putative high affinity and forming selective interactions with the ARTD-1 binding site not available to be formed with ARTD-2;
- Generate a homology model of human macrodomain-containing protein 2 (MacroD2), to identify residues in MacroD2 putatively forming interactions with analogues of a newly-identified endogenous ligand and to detect stable water molecules in the MacroD2 binding site as observed in molecular dynamics trajectories of the explicitly solvated protein-ligand complex.

Further, *in vivo* methods were used to determine the values of pharmacokinetic parameters and to establish tumor growth inhibition efficacy of small-molecule inhibitors of multiple tyrosine kinases identified as cytostatic on a subset of immortalised human cancer cell lines.

ZUSAMMENFASSUNG

In-silico-Methoden können angewandt werden, um experimentelles Drug Design zu komplementieren oder teilweise zu ersetzen. In dieser Doktorarbeit wurden computergestützte Methoden verwendet um:

- physikalisch-chemische Unterschiede zwischen den Bindungsstellen ähnlicher Proteine - dem humanen ARTD-1 und ARTD-2 - zu bestimmen, wobei die Dynamik dieser Unterschiede (aus Molekulardynamiksimulationen) berücksichtigt wurde;

- mit Molekulardynamiksimulationen energetisch günstige Proteinkonformationen von ARTD-1 zu bestimmen. Diese Proteinkonformationen können unter Berücksichtigung der zuvor berechneten physikalisch-chemische Unterschiede zwischen ARTD-1 und ARTD-2 für die Anwendung in *in-silico*-Docking weiter selektiert werden;

- eine Scoring-Funktion für die Bewertung kleiner Moleküle an der ARTD-1-Bindungsstelle zu definieren, mit dem Ziel, hoch-affine Binder aus eine *in-silico* Bibliothek zu identifizieren;

- neue Moleküle, die ARTD-1 aber nicht ARDT-2 binden sollten, auszuwählen, um diese dann experimentelle validieren zu können;

- ein Homologie-Modell des menschlichen Makrodomain-Proteins 2 (MacroD2) zu erstellen; um Reste in MacroD2 zu identifizieren, die wahrscheinlich mit Analogen eines neu bestimmten endogenen Liganden interagieren, und um stabile Wassermoleküle an der MacroD2-Bindungsstelle zu ermitteln, wie sie in Molekulardynamik-Trajektorien des Protein-Ligand-Komplexes (in wässriger Umgebung) beobachtet wurden.

Weiters, Pharmakokinetische Parameter von Tyrosin-Kinase-Inhibitoren, sowie deren Potential Tumorwachstum zu hemmen, wurden in-vivo bestimmt.

1. INTRODUCTION

1.1. DRUG DEVELOPMENT PROCESS

This introduction navigates from large to small scales along the drug development process. This is undertaken prior to delving into the application of *in silico* (computational) atomic-level studies to a drug development goal, to define the achievable scope and limitations of these studies relative to a human-level objective.

It is not possible to define a unique development process leading to a drug for human use. Until very recently since the advent of disease treatment by xenobiotics, compounds presenting therapeutic benefits were largely identified from natural products based on observations following their administration to patients made outside of a controlled trial situation, and without understanding their mechanisms of action. These natural products originated from organisms having sometimes evolved to produce compounds in response to a similar threat [1]; through human experimentation, such compounds occasioned to meet the right indication enough to be identified as a treatment.

This process has significantly evolved in line with molecular-level gains in understanding disease mechanisms. In recent decades, identification of a drug target, an aberrantly acting or expressed biomolecule (or its upstream/downstream partners) has defined the first step of projects aiming to identify compounds modulating this aberrancy. Many such targets have been proteins with a role in cellular signal propagation, whose over-signaling is associated, for example, with amplified growth rates characteristic of cancerous cells. Such so-called "translational research" projects (depicted in Fig. 1) follow the selection of the protein target by a campaign to find small-molecule compounds ("hits") binding this target with high potency and preventing it from overactive function, often by excluding the proteins' endogenous substrate [2]. In most cases, these hit compounds are then modified ("derivatised") to optimise potency and the selectivity of binding to the intended over other targets. Later derivatives meet further downstream requirements: resistance to cellular efflux, aqueous solubility, cell permeability, metabolic stability, minimal toxicity and most importantly, measurable modification in disease phenotype in a model system (cellular or animal) thought to be reflective of an expected phenotypic change at the level of the human organism.

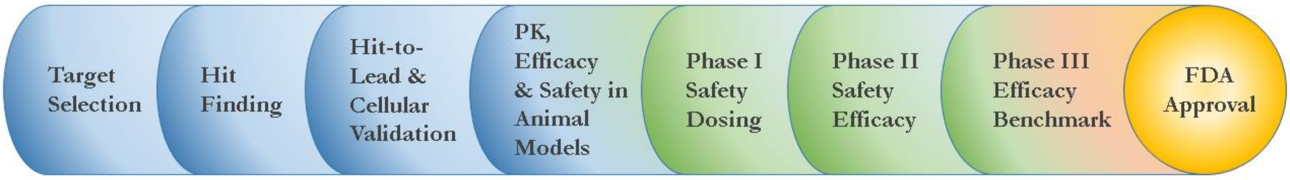


Figure 1. Translational research-based drug development process

If the preclinical evidence collected in model systems is persuasive, clinical development in humans may follow, commencing with minimal toxicity dose range determinations and proceeding to efficacy studies in groups counting dozens and then hundreds of patients, with comparisons to conventional therapy. Compelling results and the economic potential the drug compound may see it become a standard treatment.

In silico studies have potential roles to play throughout this process (the reader is directed to [3]). In the following, focus is lent to their preclinical usage, notably to identify compounds binding with high affinity to a preselected target. The discussion includes mention of steps upstream and downstream from this identification step, as their outcomes may influence the success of computational undertakings.

1.1.1 TARGET IDENTIFICATION

A therapeutic target is a biomolecule whose activity has the potential to be modified by an applied (generally xenobiotic) substance resulting in a desired phenotypic change at the level of the cell and, ultimately, organism. The instant discussion will center on the protein class of therapeutic targets, and more particularly on enzymes, which catalyze the conversion endogenous substrates to products.

The enzymatic proteins of this thesis convert small-molecule substrates to products covalently bonded to recipient proteins, or act to remove these. Such products constitute a post-translational modification (PTM) of the recipient; these PTMs can serve to direct the modified protein to perform a new function affecting cellular phenotype [4]. Over-activity of enzymatic proteins may result in modification of cellular phenotype to an extent that it contributes to or drives disease. Indeed, the definition of a therapeutic target includes both disease phenotype “drivers” (whose modulation may resolve the disease) and “passengers” (whose modulation may alleviate disease symptoms).

Two approaches have principally been used to identify potential drug targets. In so-called “top-down” approaches, all cellular processes and interaction networks known to be solicited in generating a disease phenotype constitute the starting set from which a target is identified. By contrast, “bottom-up” approaches examine common genetic or proteomic variants observed in the context of a disease to identify putative targets [5].

A potential target is preliminarily validated (generally by loss- or gain-of-function studies [6, 7]) in model systems, such as a cell or animal. A major hurdle is then to define a suitable model that recapitulates the human disease mechanism, its role in generating a measurable phenotype (or a marker of this phenotype), and extent-appropriate changes in phenotype on perturbations of the system.

Implication in disease phenotype is not sufficient to define a therapeutic target. This target must further be capable of interaction with a xenobiotic such that its activity is modulated in the direction of restored health. Proteins with such a capacity are called “druggable” [8]. Principally, these xenobiotics are in the form of a small-molecule compound or antibody. In the case of small-molecule drugs, binding is usually to a cavity on the protein surface, and facilitated by electrostatic, π-orbital stacking and van der Waals interactions.

Identification of novel therapeutic targets is clearly challenged by the use of abstracted target validation models and the need for druggability prediction; moreover, there is a risk that the target acts in a redundant cellular process to generate the phenotype, a fact which may only later become apparent when the true biological system circumvents the drug-driven change by activating redundant pathways to generate the disease phenotype. As per Medina-Franco, “[d]rug design at the single molecular target level ‘is blind’ to other processes that are inevitably connected through complex networks with higher levels of the hierarchical nature of biological systems” [9].

Therefore, evaluation of outcomes of a drug development campaign should include a review of its shortcomings. Failure to identify the correct target or indeed targets that must be (simultaneously) modulated to modify a disease phenotype does not imply failure in the approach of designing target-selective drugs [10]. The future may well lie in the use of drug cocktails composed of potent therapeutics each modulating its own validated target among a target set.

1.1.2 SMALL-MOLECULE HIT COMPOUND IDENTIFICATION

On the assumption that a valid therapeutic target has been identified, the first objective according to Fig. 1 is to find a series of small-molecule “hit” compounds binding a specified region of the target with high affinity such that it brings about the desired change in target protein function. We consider the case of a xenobiotic small-molecule ligand reversibly binding an enzymatic protein at the binding pocket of the endogenous substrate. By excluding the substrate, this competitive ligand decreases the rate of formation of the endogenous product; the function of this enzymatic protein is thereby effectively modulated.

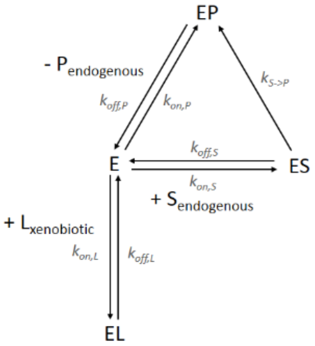


Figure 2. Enzyme kinetics relative to an endogenous substrate, product and xenobiotic ligand

With reference to Fig. 2, which assumes negligible return of product to substrate, this modulation can be discussed quantitatively. In absence of ligand (L) and assuming a non-limiting rate of product release from the complex, the enzymatic function of the protein at equilibrium is characterised by the rate of generation of product, v.

$$v = d\left[P\right]/dt = k_{cat}\left[ES\right]$$

The first step of the reaction (substrate binding) is assumed to reach equilibrium such that the ratio of reactants to products is constant:

$$K_S = \frac{[E][S]}{[ES]}$$

Product generation per unit time is then proportional to the concentration of "free" enzyme, $[E]$ available to act on the endogenous substrate. As per Fig. 2, the xenobiotic ligand effectively sequesters the enzyme, lowering the concentration of free enzyme available to the endogenous substrate. The free enzyme in presence of the xenobiotic ligand is then $[E]-[EL]$.

We define K_i of the xenobiotic ligand ("i" for inhibitor of enzymatic function) at equilibrium as:

$$K_i = \frac{k_{off,L}}{k_{on,L}} = \frac{[E][L]}{[EL]}$$

To decrease the value of v, the fraction of enzyme complexed by the ligand at equilibrium can be maximised by two routes: increasing the ligand relative to enzyme and substrate concentrations, and by decreasing the value of K_i, a measure of the inherent binding affinity of the ligand for the enzyme. Because of the risk that xenobiotics generate an unwanted and deleterious ("toxic") phenotype or have metabolites doing the same (this risk often being heterogeneous in a patient population and difficult to predict) [11], it is generally an objective to limit the amount of ligand administered to patients. Accordingly, the focus is turned to identifying compounds having a high binding affinity (low K_i) for the target. It is to be remarked, however, that compounds established as non-toxic over a wide dose range do not need as high a binding affinity for their target as generally sought for novel xenobiotics. This motivates the repurposing of drugs having established safety in humans a modulators of different targets.

The value of K_i can be obtained experimentally; the rate of product generation (or substrate disappearance) is measured at fixed total enzyme concentration and varying substrate concentrations, and plotted as Fig. 3.

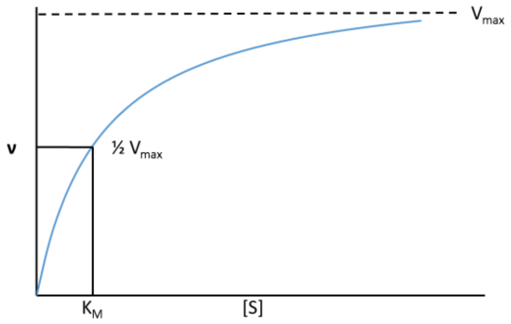


Figure 3. Reaction speed as a function of substrate concentration in the absence and presence of a substrate-competitive, reversibly-binding ligand

At low substrate concentration, the reaction rate varies near linearly with $[S]$. The asymptote corresponds to the maximum velocity, $V_{\max}$, when all enzymes substrate-bound. On reaching $V_{\max}$, the rate of enzymatic activity is no longer limited by recruitment of substrate to the enzymatic binding site; it only depends on the rate of conversion of enzyme bound ligand to product k_{cat}. This is described by the Michaelis-Menten equation:

$$v = V_{max} \frac{[S]}{[S]+K_M}$$

where $K_M = \dfrac{k_{off,S} + k_{cat}}{k_{on,S}}$

$$V_{\max} = k_{cat}[E]_0$$

K_M is the Michaelis-Menten constant. It corresponds to the substrate concentration at which $\frac{1}{2}V_{\max}$ is reached. Accordingly, a small value of K_M indicates that only small amounts of substrate are required to saturate the enzyme. Although K_M is obtained from the above plot generated at a given enzyme concentration, it is invariant to this concentration and enzyme-substrate pairs has a single characteristic K_M (at a given pH and temperature). Interestingly, physiological substrate concentrations are often close to their target enzyme K_M value; near to K_M, small changes in substrate concentration translate in large effects on enzymatic rate, attesting to the finely-tuned response systems that enzymes represent.

K_i is determined by the same assay as K_M, but in the presence of a set ligand concentration. Recasting the Michaelis-Menten equation, one obtains a Lineweaver-Burk Plot in presence and absence of a ligand.

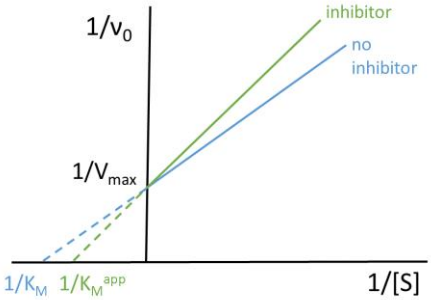


Figure 4. Lineweaver-Burk plot to determine K_M and K_M^{app}

As per the above figure, the presence of ligand leads to the measure of an "apparent" K_M (K_M^{app}), differing from that obtained in absence of ligand. The enzyme-substrate K_M can be factored out of K_M^{app} as follows to obtain the value of K_i.

$$K_M^{app} = \alpha K_M$$

$$\textit{where } \alpha = 1 + \frac{[I]}{K_i}$$

Like K_M, K_i is invariant to substrate and enzyme concentrations used in its determination. Provided then that experiments are run at the same pH and temperature (and indeed these are often set to match their narrow physiological values), K_i values of different ligands can in principle be compared to one another (nonetheless a discussion of the errors in their measurement follows). Alternatively, enzyme inhibitors can be characterised by their inhibition concentration (IC_{50}, at which enzyme rate is reduced to 50% of $V_{\max}$) determined in experiments fixing both enzyme and substrate concentrations, while varying the concentration of inhibitor. The IC_{50} value obtained depends on the substrate concentration employed, and K_i and IC_{50} are related by the Cheng-Prusoff equation.

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_M}}$$

where $[S]$ is the substrate concentration of IC_{50} determination and K_M the constant of that substrate-enzyme pair. IC_{50} values are often determined at substrate concentrations near K_M, where the change in rate per unit of substrate concentration is largest. Indeed, IC_{50} values collected from a large database relative to different proteins and measured by different research groups were found to differ from K_i values by a factor of 2 to 2.3 [12], attesting to the general usage of such substrate concentrations by experimentalists.

It is worth underlining here that whereas Fig. 3-4 refer to experimental outcomes, Fig. 2 is presented as an underlying model. Already, we note that this model is highly simplified. Proteins are dynamic structures that explore a set of conformations within the environment of the human body [13]; protein-ligand complexes often present an adapted conformation allowing enthalpically-favourable intermolecular interactions [14]. If the set of protein conformations observed in the complex are not among those frequently adopted by the apo protein, there may be a time- or energetic-barrier to adopting this new conformation. As such, the model of complex formation might be adapted as in Fig. 5, which presents two theories of protein conformation adaptation relative to protein-ligand complex discussed in 1.2.1.

From the foregoing discussion, the reader should retain that high binding affinities (low K_i or IC_{50}) can be quantified and are desired to limit the unknown risk of ligand-associated toxicities. As we move to *in silico* methods to identify putative high-affinity ligands, we have already alluded to a challenge in doing so; effective screening for ligands will have to take into account possible conformational variants of the target protein.

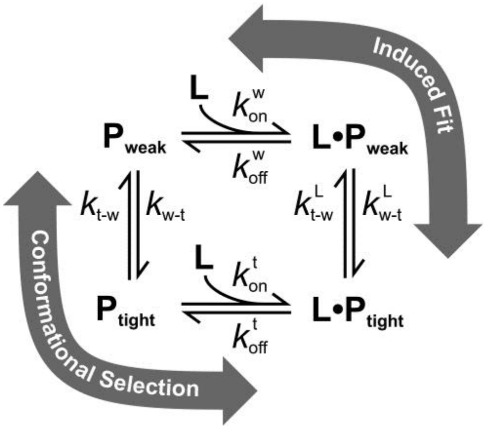


Figure 5. Enzyme kinetics relative to a xenobiotic ligand with conformational variation of the enzyme, from [15]

1.1.3 SMALL-MOLECULE HIT-TO-LEAD DERIVATISATION

Following hit compound identification, addition or modification of its chemical moieties may be undertaken to improve IC_{50} or K_i values. To allow for such modifications while limiting molecular weight, it is advantageous for hits to have a high ligand efficiency (K_i normalized by ligand heavy atom count or molecular weight) [16]. A crystal or NMR structure of the hit complexed with its target protein is helpful to identify where modifications or additions can be made. The selectivity, characterised by a series of ratios of ligand K_i of each unintended over intended target, may also be improved at this stage by modification leading to selective interactions that are unique to the target binding site (the reader is directed to [17]). Additionally, changes may be made to affect aqueous solubility [18], and to replace moieties known to be toxic [19] or metabolised [20].

Once a compound with sufficient potency and selectivity is found, it will generally be evaluated in a model system of the disease. In the case of cancer, this model system may be immortalised or patient-derived tumor cells growing in a dish. The latter can offer the advantage of more closely reflecting their biochemical makeup of origin (though both will differ due to culturing outside of the tumor environment; Domcke [21], for example, reports on differences in gene copy number, mutations and mRNA expression profiles of ovarian cancer cell-lines relative to patient tumor samples). In analogy to IC_{50} determinations on the target enzyme, measures of cell viability (such as by MTT assay, [22]) over a dose range of compound provide a concentration causing 50% of cells to lose (metabolic) viability.

1.1.4 SMALL-MOLECULE COMPOUND *IN VIVO* EVALUATION

Despite the successful identification of small-molecule compounds binding their target with high affinity and demonstrating activity in cellular disease models, compounds often lack efficacy in animal models ("*in vivo*"). *In vivo* efficacy requires that (a) sufficient compound reach its target biomolecule and (b) that this biomolecule remains critical to the extent of disease phenotype in the *in vivo* environment.

The preference for oral over caregiver-administered drug delivery carries a risk of systemic drug concentrations being limited by inefficient absorption through the intestinal wall. On reaching the systemic circulatory system, drug is distributed throughout the body and to be effective, it must reach its target location; the extent of distribution among the compartments of the body will often depend on the lipophilic and hydrophilic properties of the compound. The drug may be metabolised into inactive products, reducing the effective drug concentration (though metabolic pathways have been used to the advantage of the prodrug strategy), and/or excreted from the body before bringing about the desired phenotypic change. Insights into ADME (Absorption, Distribution, Metabolism and Excretion) of a xenobiotic are often gained through evaluation of pharmacokinetic parameters in animal models, in part based on metrics derived from the evolution of xenobiotic concentration in blood plasma over time.

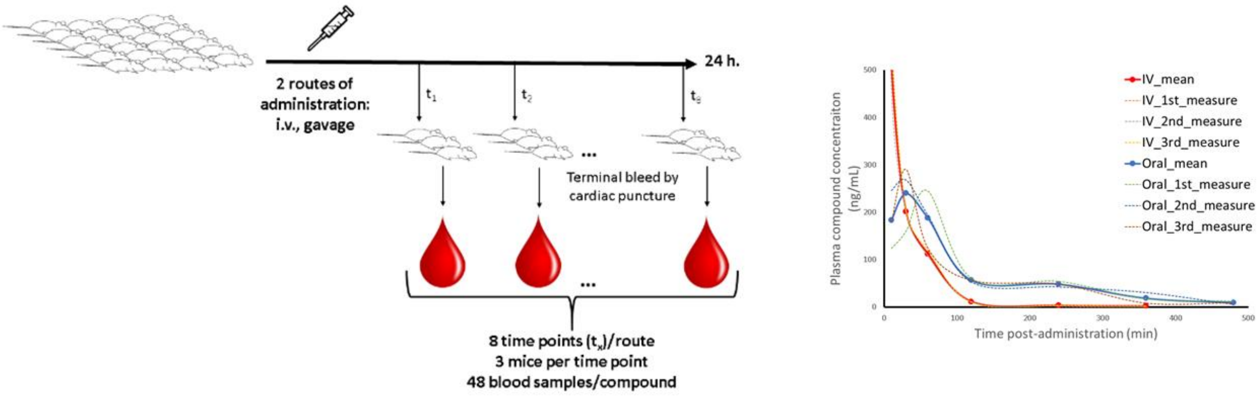


Figure 6. Pharmacokinetic evaluation in mice relative to compound 11d of [23]
Right-hand panel pictorially represents the study design leading to the curve of compound concentration in blood plasma over time given in the left-hand panel.

Useful metrics include the area-under-the-curve (AUC) of blood plasma compound concentration over time, from which a number of further parameters are calculated, such as oral bioavailability ($F\%$, fraction of orally administered compound reaching the systemic circulation unchanged), compound elimination half-life ($t½$) and maximum compound concentration following oral exposure (C_{max}).

The second efficacy requirement is that the target remain relevant to the extent of disease phenotype *in vivo*. Cellular networks and processes may be differently established and activated by the three-dimensional vascularised environment of multicellular organisms. To this end, relatively poor model systems exist to recapitulate the future human environment for efficacy evaluations [24,25]. Nevertheless, animal model studies are useful in that they better reflect the drug-response of a disease than a dish of cells. For this reason, regulatory authorities will often require animal models studies, which in the case of cancer, are often partly based on the measure of tumor growth inhibition in a mouse implanted with human tumor cells.

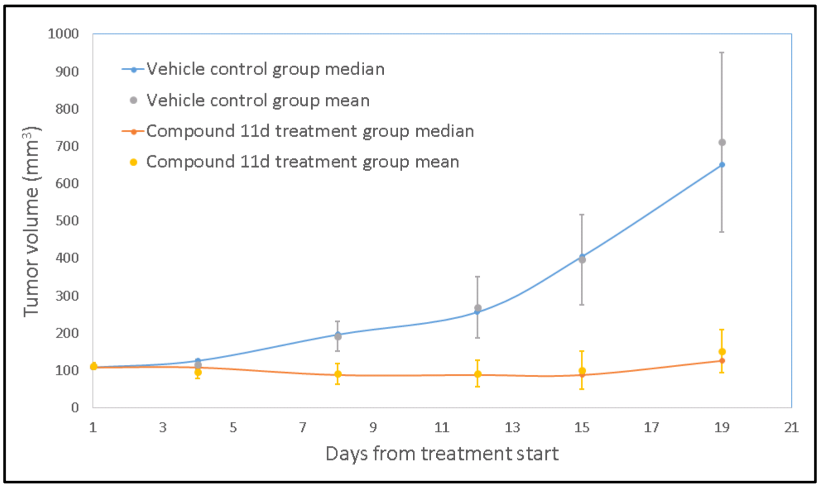


Figure 7. Tumor growth in a xenograft mouse model treated with a therapeutic compound (11d) and vehicle-control [23]

Lastly, *in vivo* studies often provide the first evaluation of toxicity: the negative effects of xenobiotic compound (and their metabolites) binding to target and non-target biomolecules in a diversity of healthy cells. This merits a brief word on the identification of targets in view of not efficacy, but toxicity: in some cases, a disease phenotype is associated with overexpression of a given biomolecule. If returning biomolecule concentrations to normal levels leads to a decrease in extent of the undesired phenotype, this biomolecule may be considered as a candidate for targeted therapy. Yet it is important to ensure that this target is not expressed and required for cellular viability within all or even a subset of healthy cells critical to well-being. Indeed, if cellular uptake is assumed comparable over all cells, it is clear that a greater fraction of normally expressed target in healthy cells will be sequestered at a given concentration of ligand than overexpressed target in disease-causing cells. The consequent extent of healthy cells death may outweigh the benefits of returning diseased cells to a healthier phenotype. Yet in line with this concept, a recent publication [26] has underlined a promising new approach aiming at the “Achilles heel” of cancers: targeting essential gene products that are uniquely under-expressed in cancer cells.

1.2 APPLICATION OF *IN SILICO* METHODS TO PRECLINICAL DRUG DEVELOPMENT

If *in silico* methods are used in drug design, they must be equivalent to or advantageous over available experimental methods. Such advantages might reside in:

- Time- or cost-efficiency in achieving a given goal;
- Capacity to achieve a given goal.

High-throughput virtual screening was used in this thesis relative to two objectives:

- To define a scoring function to apply to small-molecule compound poses within the ARTD-1 binding site with the aim of enriching the fraction of high binding affinity compounds of a library above a selected score cut-off;
- To propose for experimental validation small-molecule compounds binding ARTD-1 with putative high affinity and forming selective interactions with the ARTD-1 binding site not available to be formed with ARTD-2.

The following discussion considers the search for small-molecule drugs competing with the endogenous substrate to reversibly bind an enzyme. As in the experiments discussed previously, *in silico* approaches “expose” the target protein to a compound and a signal is read as a measure of binding affinity. The wet lab experiment requires no knowledge of the conformational ensemble of the protein-ligand complex established in the test environment; it does require a physical sample of each member of the library for testing. By contrast, no physical samples are needed for *in silico* evaluations, underlining their interest for large libraries and highly novel compounds. Knowledge of protein conformation(s) may, however, be determinant to *in silico* success.

To identify potent ligands from a ligand library, a set of conformations (“poses”) of each ligand in the target site of the protein is generated and must include at least one representative of the true ensemble of conformations adopted by the corresponding protein-ligand complex. This or these representative(s) then need to be extracted from the generated set and ranked relative to representative poses of the protein with the remaining ligands of the library. This ranking should distinguish ligands having high binding affinity from the rest.

If a high-performing combination of docking algorithm and ranking (“scoring”) function can be found, *in silico* screening offers advantages of speed and diversity of evaluated compounds over wet lab methods. Prior to any compound synthesis or evaluation, complexes may be selected or eliminated based on their particular protein-ligand interactions, offering the possibility to favour compounds forming selective interactions that are unique to the target, and to avoid interactions with residues known to undergo drug-induced resistance mutations *in vivo*.

1.2.1 DOCKING

The first requirement, the generation of an ensemble of ligand poses in the protein binding site that includes a representative of the true complex, is partly an issue of algorithms. In general, the conformation of the protein is approximated as fixed and originates from crystal, NMR, homology-modelled or molecular dynamics-generated structures. For the purpose of speed, the only degrees of freedom during docking are rotations about ligand dihedral angles. Retained poses are those in which the ligand “fits” the protein binding site without excessive steric clashes between their atoms represented as spheres. But even an exhaustive search of ligand conformations within the binding site will not generate a representative of the true complex if the rigid protein structure is incorrect. As per McCammon [27]: “Virtual screening results are influenced by the PDB structure chosen for screening because the fit of individual ligands in the pocket is affected by even the most minor structural changes”.

Indeed, proteins are not rigid structures, and they often present alternative conformations to their apo structure when complexed with a ligand, as evidence by crystal and NMR investigations. To partly account for these alternative conformations, the docking campaign can include a set of rigid conformations of the protein (for “ensemble” or “conformational docking”). This invites the question of how to generate and select this set.

While experimental evidence lays plain the existence of differing protein conformations in apo and complexed structures, they have not conclusively revealed the process of this conformational change. Accordingly, hypothetical models to fit experimental observations have been proposed and currently at least two are in circulation: in the “induced fit” mechanism [28], the ligand binds a protein conformation from among an ensemble of protein conformations that does not include that of the complex. Subsequent to or concurrent with ligand binding, the protein “responds” to the ligand by effecting a conformational change to reach a new state maximising the favourable free energy change of complex formation. By contrast, conformational selection hypothesizes that the ligand “selects” (binds to) a protein conformation that pre-exists within the apo protein conformational ensemble. As per Valente [29]: “In the free state, protein regions displaying conformational diversity exhibit equilibria among pre-existing conformations. In the presence of a ligand, one of these conformations is stabilized, so that the ligand does not need to induce a new conformation. Upon ligand binding there is a population shift toward the bound conformational state. Conformational diversity of binding sites of several proteins has been measured and has important practical as well as thermodynamical consequences: binding sites can be mapped without prior knowledge of the ligand and also evolution of binding sites depends mostly on the free state, occurring at least partially independently of the ligand”. The last part of this concluding sentence hints at a third hypothesis: that the process involves a mixture of both mechanisms, with ligand selection of a protein conformer "near to" that of the complex and some mild structural adaptations of the protein (and ligand) thereafter.

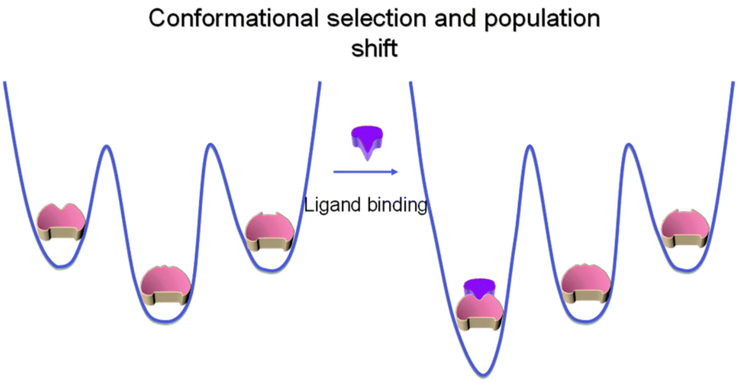


Figure 8. An overview of the free energy landscape, conformational selection and population shift from [30].

For the purpose of *in silico* docking to rigid protein structures, the conformational selection hypothesis invites alternative apo protein conformations be taken into account and indeed hints at the use of molecular dynamics techniques to generate a conformational ensemble. But which ensemble members should be included in the docking campaign? One might select a set of MD-generated protein conformations based on maximising their structural diversity. Yet this could lead to the inclusion of protein conformations corresponding to rare, high-energy states.

This is intuitively (though not definitively, see [31]) unsatisfying because although ligand binding may change the energy landscape of a protein to stabilise a given protein conformation, it must be to an extent that the free energy of the complex is lower than that of the free protein and ligand, as well as the free energy of the protein-endogenous substrate complex and all alternative protein-ligand complexes: commencing from a high free energy protein conformation implies a significant stabilization through ligand binding is necessary. Though this is not impossible, it suggests a rare ligand binding with exceptional binding affinity.

For this reason, a diversity of states among those of the highly-populated free energy basins, and not barriers, of MD simulations of the apo protein were selected for the docking projects presented herein. This position is supported by the following excerpt [32] regarding two targeted small-molecule therapeutics, Dasatinib and Imatinib: "Dasatinib is ~350 times more potent an inhibitor of the Abl kinase than imatinib. Dasatinib recognizes the active conformation of the kinase domain, which is likely to be the predominant form in cancer cells. One reason, therefore, for the higher affinity of dasatinib for Abl is due to its not having to stabilize a less populated conformation."

1.2.2 SCORING

Following docking and the generation of sets of protein-ligand complexes presumed to include a representative of their true conformation, it is necessary to extract this representative and compare it with representative complexes of other ligands according to an estimation of binding affinity.

According to thermodynamics, spontaneous complex formation from its separate parts occurs with a decrease in free energy ($\Delta G_{binding} < 0$). Consistent with this for complex formation at equilibrium:

$$\Delta G_{binding,eq} = -RT\ln(K_i)$$

where $R = 8.314\, {}^{J}\!/_{mol\cdot K}$

As K_i decreases and the ligand binds the protein with higher affinity, $\Delta G_{binding}$ becomes more negative and complex formation more favorable. For completeness, it is mentioned here that $\Delta G_{binding}$ or $\Delta(\Delta G_{binding})$ between two complexes with two (preferably similar) ligands may be evaluated by molecular dynamics-based techniques of thermodynamic integration and free energy perturbation. Both are computationally demanding and hence incompatible with evaluations of complexes formed with large ligand libraries. We therefore need an estimator of $\Delta G_{binding}$ ($\Delta \hat{G}_{binding}$); this estimator can include any term and its equation can take on any form, provided that it correctly reproduces experimental values.

Decomposition of free energy

Free energy changes can be decomposed into contributions of enthalpic and entropic changes on binding:

$$\Delta G_{binding} = \Delta H_{binding} - T\Delta S_{binding}$$

Changes in entropy of the solute on binding are presumed to be both unfavorable to $\Delta G_{binding}$, consequent to restriction of ligand and protein conformational freedom. In contrast the release of structured water from the binding site and ligand is favourable. The extent of entropy change is challenging to estimate relative to a static picture of system components: nonetheless, we note that the conformational entropy change on binding tends to be unfavorable with increasing ligand rotatable bonds (due to loss of torsional entropy), and that decreases in solvent accessible surface area (SASA) of both the ligand and protein with binding include liberation of structured water surrounding hydrophobic regions, increasing $\Delta S_{binding}$.

Intramolecular protein-ligand interactions consistent with hydrogen bonds, salt bridges, van der Waals interactions and π-orbitals stacking, among others, are apparently throughout crystal and NMR structures of complexes. Their formation is exothermic and (to the extent that they are more favorable with each other than with water) they decrease system enthalpy, $\Delta H_{binding}$, on binding (the reader is directed to the following perspective article on molecular interactions

[33]). Many such interactions can be modeled as a function of inter-atom distance by equations of charge (Coulomb) and van der Waals (Lennard-Jones) interactions, as presented in the empirical potential energy forcefield discussion hereafter.

Linear interaction energy

The above discussion provides one possible selection of terms for $\Delta \hat{G}_{binding}$. In building a mathematical model, which terms should be included and how should they be combined? One possibility issues from the semi-empirical LIECE (Linear Interaction Energy (LIE)[34] with Continuum Electrostatics) method [35], which is based on the linear combination of van der Waals and electrostatic terms (evaluated using an implicit solvent model and calculated with CHARMM using the Poisson equation solved numerically by the finite-difference method for determination of the electrostatic component of solvation energy [36,37]). Each term is evaluated as the difference between end points of the thermodynamic cycle of ligand binding, i.e. the bound and free states [38], and weighted to fit experimentally-derived values of $\Delta G_{binding}$.

Solvation effects: Finite difference Poisson equation

As the electrostatic component of solvation energy was also considered in the model of a project of this thesis, a brief explanation of the Poisson equation and its resolution is provided here. In a continuum solvent model, all solute atoms are considered explicitly as point-charged particles having a low dielectric constant (i.e. low ability to polarize in response to an external electric field; generally in the range of 2-4 for proteins and small-molecules), while the solvent is implicit and represented by its dielectric constant (~80 in bulk water, where a high value reflects a high ability to re-orient water molecule dipoles and thus optimise interaction with an electric field). The electrostatic potential at a given point in space (φ(r)) is dependent on solute's charge density (ρ(r)) and the local dielectric (ε(r)) according to the Poisson equation:

$$\nabla\left[\varepsilon(r)\nabla\varphi(r)\right] = -4\pi\rho(r)$$

On immersion of a solute (here protein or small-molecule) in solvent, induced charges appear at their boundary due to the difference between solute and solvent dielectrics, locally modifying the dielectric from the bulk values. The electrostatic potential is generally determined relative to points on a grid applied to the solute-solvent system (where each grid point is assigned a charge and dielectric based on those in proximity) according to the finite difference method. The electrostatic free energy of a given system is then obtained by integrating over the charge density and electrostatic potential values over all points. The electrostatic component of the free energy of ligand and protein binding can then be determined according to the following cycle, including the energy of transfer of each solute (protein, ligand, complex) from vacuum (or a low dielectric reference environment) to a high dielectric solvent, and the electrostatic energy of protein-ligand binding ($E_{Coulomb}$ in a medium with fixed ε) calculated according to the Coulomb equation.

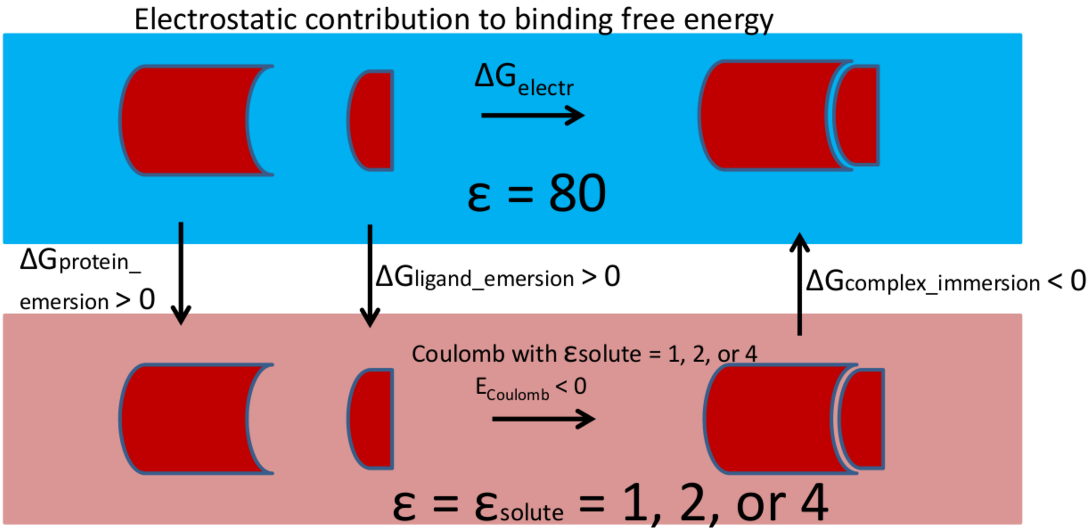


Figure 9. Thermodynamic cycle for determination of electrostatic contribution to binding free energy including solvation (ΔG_{electr})

Model building by multivariate linear regression

The values of user-selected X_i terms (such as ΔG_{electr} described above) of the protein-ligand complexes in combination with estimates of $\Delta G_{binding}$ from experimental measures form a set of equation as below allowing the fitting of the unknown coefficients (ci).

$$\Delta G_{binding} = c_1 X_1 + c_2 X_2 + ...c_i X_i$$

Building such a model presupposes two things: that the terms (X_i) employed to build the model are consistently evaluated relative to a correct representative of each protein-ligand complex, and that the error in the corresponding experimental values used to calibrate the weighting terms is low enough to capture a trend. The training set $\Delta G_{binding}$ values should ideally be uniformly distributed over the range of interest; the resultant model is used for interpolation, and not extrapolation.

The values of $\Delta G_{binding}$ used to build the model in this thesis were calculated from literature-reported K_i and IC_{50} values, the latter being more experimental condition-dependant (substrate concentration, pH, temperature) than the former (pH, temperature) but more often reported. In spite of their relative independence to experimental conditions, even K_i values suffer from variation (due to systematic errors between data sources and precision errors). Kalliokoski has estimated the expected error in K_i values of a given inhibitor-protein pair reported by different labs in the ChEMBL database [39]. IC_{50} variability between measurements has been reported by the same author [12].

Any inaccuracy in the $\Delta G_{binding}$ values used to build the model limits the maximum correlation attainable. Given this, it is worthwhile to further evaluate the model before use. Since the number of complexes having a reported K_i or IC_{50}

value and a known or highly probably binding mode in the target is limited, there is an interest to use all data in the model “training set”. No test set then remains to evaluate the models predictive performance. In such a case, a cross-validation correlation coefficient may be helpful; it is generated by leaving one data point out (“leave-on-out”) and comparing the experimental to predicted (according to a model generated by the remaining data points) free energies of the excluded point. As per Golbraikh, the cross-validation correlation coefficient is useful to eliminate poor models, but the standard cut-off of $q^2 > 0.6$ is insufficient to validate a good one [40].

Another approach to evaluate a model is to compare its correlation and cross-validation correlation coefficients with those of a large number of models generated by datasets in which each set of X_i values is associated with a randomly-generated value of $\Delta G_{binding}$. If the proposed model is superior in both coefficients to those on random $\Delta G_{binding}$ assignment, the model outperforms a random selection.

Nonetheless, even having passed various tests to eliminate clearly poor performing models, the remaining model may suffer inaccuracies and accordingly the output values (“scores”) should not be considered as absolute measures of binding affinities but rather as coarse indicators of a range in which binding affinity is likely to lie. It is then reasonable to define a “score” cut-off above which true high affinity binders (“true positives”) are likely to significantly outnumber false positives. To this end, a receiver operator characteristic (ROC) plot may be helpful in selecting this cut-off value.

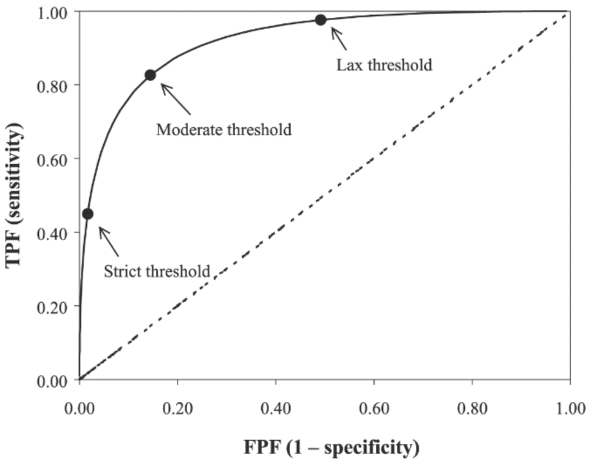


Figure 10. Representative ROC plot from [41]

If a sufficient number of ligands remain above the cut-off, a diverse set of these can be selected for experimental evaluation.

1.2.3 DIGITAL LIBRARIES

Once a docking algorithm and a target-calibrated scoring function are defined, the combination may be applied prospectively to identify putative high affinity binders of the target protein. This presupposes that such high affinity binders are in the library to be screened. For this reason, large libraries are of interest.

On the other hand, it is more efficient to tailor a library in advance of docking if desirable characteristics of the potential inhibitors are known (the presence of hydrogen bond donors or acceptors in view of the target binding site and further drug-like properties). Ligands containing privileged chemical moieties known to bind sub-pockets of the target binding site may help to ensure correct binding of that motif, while the remainder of that ligand explores new pockets.

In the fragment-based docking approach known as "ALTA", small fragments (M_W < 300 g/mol) with limited torsional degrees of freedom are first docked to the target binding site and ranked according to metrics of binding affinity. High ranking fragments then serve to perform a substructure search on a large digital compound library, resulting in a collection of ligands containing these select moieties. These ligands are then docked to the binding site, with conformational filtering to ensure their respective selected fragments are positioned as previously identified, prior to scoring the ligand pose. The ALTA approach may be of particular interest when the project objectives include the identification of a ligand binding strongly to a new sub-pocket of the binding site, as was a goal in this thesis.

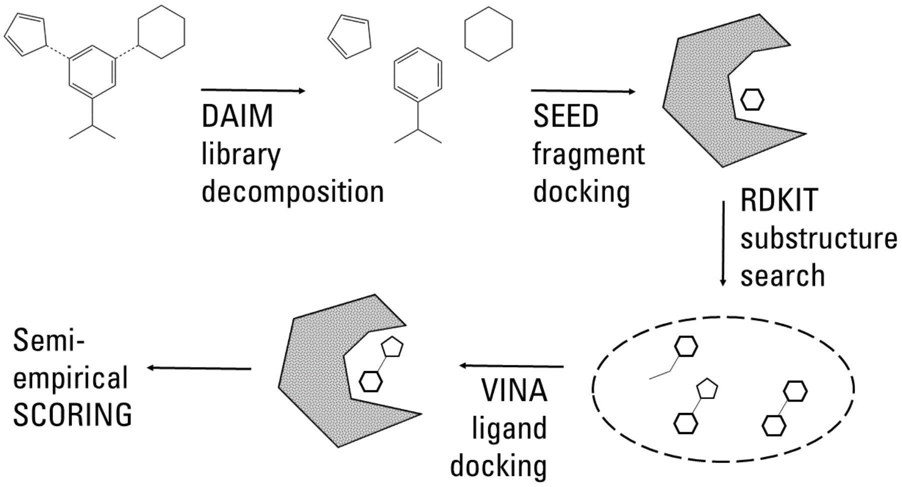


Figure 11. Fragment-based ALTA procedure as implemented in the present thesis with use of DAIM [42], SEED [43,44], RDKIT [45], VINA [46] and LIECE-type evaluation of binding affinity [35]

A last approach to library definition is mentioned here: as discussed previously, compounds having demonstrated low toxicity are of great interest to repurpose [47] because they have already overcome a major hurdle in development that is difficult to predict in advance. This motivated the docking (as of "VINA ligand docking step" in Fig. 11) of a collection of post-clinical trials phase I compounds in a project of this thesis.

1.2.4 HOMOLOGY MODELLING

Homology modelling was used in this thesis to:

- Build a homology model of human macrodomain-containing protein 2 (MacroD2) for which no crystal or NMR structure was available.

Proteins are chains defined by their particular sequence of covalently-bonded amino acid residues, of which 21 naturally occur in humans. Whereas protein amino acid sequences are widely available, their corresponding three-dimensional structures are less so. One option, then, if a structure of a protein of known sequence is required, is to build a comparative homology model of this structure. This requires that its sequence be homologous to another protein for which a structure is available. Accordingly, a first step is to perform a sequence alignment of the protein to be modelled over a large database of known sequences in order to find potential templates. Given the need for speed in view of the breadth of potential alignments, alignment algorithms are heuristic; the BLAST (Basic Local Alignment Search Tool) algorithm [48] preliminarily identifies short matching regions before aligning neighboring segments and evaluating the result. The selection of aligned protein to use as template for a homology model can be guided by the degree of sequence identity, from which expected structural differences have been quantified [49-52].

The target-template sequence alignment and template structure are then used to generate a three-dimensional model of the target structure. MODELLER [53,54] implements comparative modeling by “satisfaction of spatial restraints” derived from the sequence alignment and resulting in a set of probability distribution functions (pdfs) to constrain main and side-chain conformations. These pdfs are step-wise added to a single objective function relative to which the corresponding subset (and ultimately complete set) of modeled atomic coordinates are minimised. The lower the final objective function value, the better the fit of a given model to the applied restraints.

1.2.5 MOLECULAR DYNAMICS

Molecular dynamics was used in this thesis relative to:

- determine physico-chemical differences between the binding sites of two related proteins, human ARTD-1 and ARTD-2, taking into account the intrinsic flexibility of these enzymes as observed by molecular dynamics simulations;
- identify protein conformations falling within densely explored regions of phase space (“energetically-favourable” basins of kinetically-clustered conformations) of ARTD-1 that are distinct from crystal structures and present physico-chemical differences relative to ARTD-2 for use in a high-throughput *in silico* docking campaign;
- evaluate the stability of a homology model of human MacroD2;
- identify residues in MacroD2 putatively forming interactions with analogues of a newly-identified endogenous ligand;
- detect stable water molecules in the MacroD2 binding site as observed in molecular dynamics trajectories of the explicitly solvated protein-ligand complex.

Biological systems are systems in motion. Evidence of this issues from experiments on small and large constituents of cellular systems: from IR-observed bond vibrations, to protein crystallographic B-factors, to atomic fluctuations derived from NMR relaxation rates and structural ensembles, to temporal evolution of distance from FRET measurements [55-58]. To the extent that experimental techniques, though quantifying a metric of temporal motion within a system, are limited in resolution, either by measuring values at a level of abstraction from constituent atoms or by providing only a series of static representatives of a system, molecular dynamics provides a means to evolve classical degrees of freedom of atoms in a system over time. As per [59], one use of molecular dynamics is to "obtain a description of the system at equilibrium, including structural and motional properties".

If we consider a collection of single atoms in a system and hypothesize that their individual movements can be approximated by classical laws of physics, Newton's equation applies to trace their motion over time, in accord with the forces applied to the constituents of the system.

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d^2}{dt^2} \mathbf{r_i}$$

These forces can act to limit the available values of each degree of freedom of the system. An atom of a protein in a biological system at body temperature is covalently bonded to at least one other; a potential energy field ("forcefield") can be defined from which this force as a function of bond length is obtained from its negative gradient.

$$\mathbf{F}_i = -\nabla E(\mathbf{r}_i)$$

Ideally, this field should be represented by continuous function that recapitulates experimental observables. The Morse potential is an example of an empirical forcefield model of bond potential, appropriately including a distance limit above which potential energy becomes distance-invariant and atoms are no longer bonded; when atoms approach so as to overlap their orbitals, potential energy rapidly rises, translating in the expected strong force to return the inter-atom distance to that at a potential energy minimum.

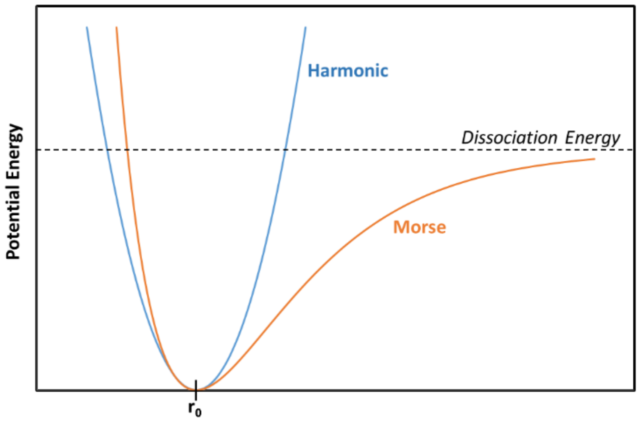


Figure 12. Morse and harmonic potentials

Given that a protein counts large numbers of bonded atom pairs, it is desirable to use a simple forcefield with few terms; in the interest of simulating biologically-relevant time scales, this is in fact essential. At human body temperature, covalent bonds do not break unassisted and the potential energy can be simplified to a harmonic potential.

$$E = K_B (r - r_0)^2$$

The above bond potential energy model appears with further terms in the widely exploited empirical CHARMM22 forcefield [60] for biomolecules as a function of system coordinates given below:

$$E(\mathbf{R}) = \sum_{bonds} K_b (r - r_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\theta (1 + \cos(n\chi - \delta)) +$$
$$\sum_{impropers} K_{imp} (\varphi - \varphi_0)^2 + \sum_{nonbond} \varepsilon \left[\left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}}$$

The reader is directed to [60] for a more complete discussion of the terms. The first five summations are relative bonded interactions (within three covalent bonds), and scale with $O(N)$. The evaluation of non-bonded terms, by contrast, are based on pairwise potentials that scale with $O(N^2)$, accounting for a large part of calculation time. For this reason, distance cut-offs are applied to limit non-bonded term evaluation. The use of a cut-off can be problematic for two reasons: (a) it implies a discontinuity at the cut-off, and (b) important long-range electrostatic interactions are ignored. The first issue can be circumvented by adding a smoothing potential over a short range at the cut-off ("SWITCH") or by shifting the potential ("SHIFT"). The second is addressed in view of the use of system replicates in all directions surrounding the simulated system in order to approximate electrostatic interactions beyond the simulation box; these replicates provide charges appearing periodically over space. This periodicity is suitable to description in Fourier space. Ewald summation (and the faster PME method) accordingly divide short-range and long-range components of electrostatic potential.

Charge interactions are determined by the Coulomb equation relative to fixed point charges centered on the atom nuclei whose magnitude and sign are either provided in the forcefield parameters or determined by distribution of formal charge as a function of relative electronegativity and the hybridisation state of the atoms; the fixed point charge representation ignores polarisation effects that might redistribute charge as a function of the environment defined beyond directly bonded partners and the generation of asymmetry dipoles (when the centroid of electronic and nuclear charge do not overlap).

Forcefields are continuously evaluated relative to experiment, offering opportunities for their improvement. In view of observations of overstabilised secondary structures attributed to forcefield bias, the CHARMM22 forcefield has since been extended by additional terms to improve treatment of backbone dihedrals (now CHARMM27) [61].

Given a defined analytical form of the forcefield, it is necessary to provide the parameters of their function. These can be determined *ab initio* by quantum mechanical calculations and/or fit to reproduce empirical data. It is important that parameters be derived in view of a given analytical forcefield form and appropriate to the system they are modeling. The CHARMM22/27 forcefield has associated parameters for all atoms (i.e. including explicit hydrogens) of proteins, lipid, nucleic acids and, importantly, water molecules and ions; these parameters are "consistently" derived, meaning that they can be used in combination with one another. Indeed, molecular dynamics simulations are often run in a "box" containing the biomolecule of interest immersed in explicit solvent (water molecules and salt ions providing physiological salt concentrations and system charge neutrality). Accordingly, the objective in setting the parameters of the non-bonded terms must "result in balanced protein-protein, protein-water and water-water interactions" [60]. The applicability of the CHARMM22 forcefield has recently been improved by the determination of parameters suitable for small-molecules [62]. Whereas prior-mentioned parameters may be accurately defined and assigned their limited breadth of atom types and environments, small-molecule are diverse and assignment of their parameters is often by analogy to a limited set of fully parameterised fragments [63].

The above discussion was undertaken to underline that potential energy forcefields may include terms that accord with physical models, but which remain approximate. Indeed, "the errors introduced by the use of empirical potentials are difficult to quantify" [59]. Nonetheless, armed with a forcefield, the equations of force above provide the basis to evolve atomic coordinates over time.

$$\frac{d^2}{dt^2} r_i = \frac{-\nabla E(r_i)}{m_i}$$

The potential energy function is derived with respect to (Cartesian coordinate-derived internal) coordinates, and evolved by its integration with respect to time. Time-reversible numerical integrators such as provided by the verlet algorithm approximate the positions, velocities and accelerations at $t+\delta t$ by a truncated Taylor series. While affected by precision limitations, what is critical is that such integrators avoid generating large instabilities and systematic drifts. The selection of the time step requires that it be smaller than the highest-frequency motion in the system. For this reason, high-vibrational frequency covalent bonds including hydrogen are often constrained to allow a larger time step to be used.

Molecular dynamics result in a trajectory (series of states) contained within phase space, where phase space consists of the set of all accessible states of a system subject to the thermodynamic constraints. The above-described system and related equations describe the evolution of the simulated system at constant total energy, which does not reflect the thermodynamic properties governing the ensemble of protein conformations in its cellular environment. We can accordingly change ensembles in which to evolve our protein by coupling the system to a heat bath and a barostat, thus moving to the statistical mechanical isothermal-isobaric NPT ensemble (where N refers to the fixed number of atoms of a closed system, which is ensured during simulation by the use of periodic boundary conditions which circulate atoms exiting a boundary the system directly to enter on the opposite boundary side). This can be achieved by reformulation of the equation of motion and its integrator to a set of equations including, at each time step, random stochastic forces (corresponding to collisions with fictive particles of the heat bath), friction and a degree of freedom corresponding to the volume of the simulation system.

With correct sampling in the NPT ensemble, the distribution of density of visited states over phase space will reflect, at equilibrium (achieved by exhaustive sampling), the relative free energies of different phase space regions: energetically favourable states are visited more often than less favorable states. Barrier regions are visited even less, though the thermostat-governed velocities will allow their occasional occupation from which different basins are reached. The density of state visitation can be clustered into (strongly connected) macrostates within a network, and a two-dimensional slice (“projection”) identified to visualize the free energy evolution between states ordered according to a given progress index. In the present thesis, a cut–base free energy profile was used to visualize and identify free energy basins, from which states were selected for a docking campaign [64]. This approach estimates the partition function (Ξ in the NPT ensemble) of a basin based on its number of visitations during the trajectory; the ensemble of states representing different transition paths between two given basins can be evaluated by the Ford-Fulkerson theorem to determine the flow rate between them [65].

1.3 AIMS AND STRUCTURE OF THIS THESIS

The main aim of the present thesis was to identify potent and selective inhibitors of epigenetic targets. Several computational methods were specifically used to:

- Determine physico-chemical differences between the binding sites of two related proteins, human ARTD-1 and ARTD-2, taking into account the intrinsic flexibility of these enzymes as observed by molecular dynamics simulations (**Chapter 2**);
- Based on molecular dynamics simulations, identify protein conformations falling within densely explored regions of phase space (“energetically-favourable” basins of kinetically clustered conformations) of ARTD-1 for use in a high-throughput *in silico* docking campaign (**Chapter 2**);
- Define a scoring function to apply to small-molecule poses within the ARTD-1 binding site with the aim of enriching the fraction of high binding affinity compounds of a library above a selected score cut-off (**Chapter 2**);
- Propose for experimental validation small molecules binding ARTD-1 with putative high affinity and forming selective interactions with the ARTD-1 binding site not available to be formed with ARTD-2 (**Chapter 2**);
- Propose for experimental validation small-molecule compounds binding ARTD-1 with putative high affinity based on a library of select post-clinical trials phase I compounds and FDA approved drugs (**Chapter 3**);
- Generate a homology model of human macrodomain-containing protein 2 (MacroD2) (**Chapter 4**);

- Identify residues in MacroD2 putatively forming interactions with analogues of a newly-identified endogenous ligand and to detect stable water molecules in the MacroD2 binding site as observed in molecular dynamics trajectories of the explicitly solvated protein-ligand complex (**Chapter 4**).

Further, studies in mice were used to determine pharmacokinetic parameters and to establish tumor growth inhibition efficacy of small-molecule inhibitors of multiple tyrosine kinases identified *in silico* (**Chapter 5**).

In a side project, a novel web-based platform was developed to render more user-friendly the suite of programs for fragment-based docking developed in the Caflisch group since 1999 (**Appendix A**).

1.4 REFERENCES

1 Ji, H. F., Li, X. J. & Zhang, H. Y. Natural products and drug discovery Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? *Embo Rep* **10**, 194-200, doi:DOI 10.1038/embor.2009.12 (2009).

2 Roses, A. D. Pharmacogenetics in drug discovery and development: a translational perspective. *Nature reviews. Drug discovery* **7**, 807-817, doi:10.1038/nrd2593 (2008).

3 Chen, L. *et al.* From laptop to benchtop to bedside: structure-based drug design on protein targets. *Current pharmaceutical design* **18**, 1217-1239 (2012).

4 Karve, T. M. & Cheema, A. K. Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease. *Journal of amino acids* **2011**, 207691, doi:10.4061/2011/207691 (2011).

5 Titov, D. V. & Liu, J. O. Identification and validation of protein targets of bioactive small molecules. *Bioorganic & medicinal chemistry* **20**, 1902-1909, doi:10.1016/j.bmc.2011.11.070 (2012).

6 Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411**, 107-110, doi:10.1038/35075114 (2001).

7 Ngo, V. N. *et al.* A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106-110, doi:10.1038/nature04687 (2006).

8 Hopkins, A. L. & Groom, C. R. The druggable genome. *Nature reviews. Drug discovery* **1**, 727-730, doi:10.1038/nrd892 (2002).

9 Medina-Franco, J. L., Yongye, A. B., Perez-Villanueva, J., Houghten, R. A. & Martinez-Mayorga, K. Multitarget structure-activity relationships characterized by activity-difference maps and consensus similarity measure. *Journal of chemical information and modeling* **51**, 2427-2439, doi:10.1021/ci200281v (2011).

10 Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nature chemical biology* **4**, 682-690, doi:10.1038/nchembio.118 (2008).

11 Kerns, E. H. & Di, L. Drug-Like Properties: Concepts, Structure Design and Methods. *Drug-Like Properties: Concepts, Structure Design and Methods*, 1-528 (2008).

12 Kalliokoski, T., Kramer, C., Vulpetti, A. & Gedeck, P. Comparability of Mixed IC50 Data - A Statistical Analysis. *Plos One* **8**, doi:ARTN e61007
DOI 10.1371/journal.pone.0061007 (2013).

13 Mccammon, J. A. & Karplus, M. The Dynamic Picture of Protein-Structure. *Accounts Chem Res* **16**, 187-193, doi:Doi 10.1021/Ar00090a001 (1983).

14 Okazaki, K. I. & Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 11182-11187, doi:DOI 10.1073/pnas.0802524105 (2008).

15 Hammes, G. G., Chang, Y. C. & Oas, T. G. Conformational selection or induced fit: A flux description of reaction mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13737-13741, doi:DOI 10.1073/pnas.0907195106 (2009).

16 Hopkins, A. L., Keseru, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nature Reviews Drug Discovery* **13**, 105-121, doi:Doi 10.1038/Nrd4163 (2014).

17 Huggins, D. J., Sherman, W. & Tidor, B. Rational Approaches to Improving Selectivity in Drug Design. *J Med Chem* **55**, 1424-1444, doi:Doi 10.1021/Jm2010332 (2012).

18 Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol* **44**, 235-249, doi:Doi 10.1016/S1056-8719(00)00107-6 (2000).

19 Smith, G. F. Designing Drugs to Avoid Toxicity. *Progr Med Chem* **50**, 1-47, doi:Doi 10.1016/B978-0-12-381290-2.00001-X (2011).

20 Fontana, E., Dansette, P. M. & Poli, S. M. Cytochrome P450 enzymes mechanism based inhibitors: Common sub-structures and reactivity. *Curr Drug Metab* **6**, 413-454, doi:Doi 10.2174/138920005774330639 (2005).

21 Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat Commun* **4**, doi:Artn 2126
Doi 10.1038/Ncomms3126 (2013).

22 Mosmann, T. Rapid Colorimetric Assay for Cellular Growth and Survival - Application to Proliferation and Cyto-Toxicity Assays. *J Immunol Methods* **65**, 55-63, doi:Doi 10.1016/0022-1759(83)90303-4 (1983).

23 Unzue, A. *et al.* Pyrrolo[3,2-b]quinoxaline derivatives as types I1/2 and II Eph tyrosine kinase inhibitors: structure-based design, synthesis, and in vivo validation. *J Med Chem* **57**, 6834-6844, doi:10.1021/jm5009242 (2014).

24 Hann, B. & Balmain, A. Building 'validated' mouse models of human cancer. *Curr Opin Cell Biol* **13**, 778-784, doi:Doi 10.1016/S0955-0674(00)00283-0 (2001).

25 Holland, E. C. *Mouse Models of Human Cancer*. (John Wiley & Sons, 2004).

26 Nijhawan, D. *et al.* Cancer Vulnerabilities Unveiled by Genomic Loss. *Cell* **150**, 842-854, doi:DOI 10.1016/j.cell.2012.07.023 (2012).

27 Nichols, S. E., Baron, R. & McCammon, J. A. On the Use of Molecular Dynamics Receptor Conformations for Virtual Screening. *Computational Drug Discovery and Design* **819**, 93-103, doi:Doi 10.1007/978-1-61779-465-0_7 (2012).

28 Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 98-104 (1958).

29 Valente, A. P., Miyamoto, C. A. & Almeida, F. C. L. Implications of protein conformational diversity for binding and development of new biological active compounds. *Curr Med Chem* **13**, 3697-3703, doi:Doi 10.2174/092986706779026147 (2006).

30 Tsai, C. J. & Nussinov, R. The free energy landscape in translational science: how can somatic mutations result in constitutive oncogenic activation? *Phys Chem Chem Phys* **16**, 6332-6341, doi:Doi 10.1039/C3cp54253j (2014).

31 Sinko, W. *et al.* Applying Molecular Dynamics Simulations to Identify Rarely Sampled Ligand-bound Conformational States of Undecaprenyl Pyrophosphate Synthase, an Antibacterial Target. *Chem Biol Drug Des* **77**, 412-420, doi:DOI 10.1111/j.1747-0285.2011.01101.x (2011).

32 Kuriyan, J., Konforti, B. & Wemmer, D. *The molecules of life : physical and chemical principles*. (New York ; London : Garland Science, Taylor & Francis Group, 2013).

33 Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J Med Chem* **53**, 5061-5084, doi:Doi 10.1021/Jm100112j (2010).

34 Aqvist, J., Medina, C. & Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein engineering* **7**, 385-391 (1994).

35 Huang, D. & Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J Med Chem* **47**, 5791-5797, doi:Doi 10.1021/Jm049726m (2004).

36 Brooks, B. R. *et al.* Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **4**, 187-217, doi:DOI 10.1002/jcc.540040211 (1983).

37 Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**, 1545-1614, doi:Doi 10.1002/Jcc.21287 (2009).

38 Huang, D. Z. & Caflisch, A. Library screening by fragment-based docking. *J Mol Recognit* **23**, 183-193, doi:Doi 10.1002/Jmr.981 (2010).

39 Kramer, C., Kalliokoski, T., Gedeck, P. & Vulpetti, A. The Experimental Uncertainty of Heterogeneous Public K-i Data. *J Med Chem* **55**, 5165-5173, doi:Doi 10.1021/Jm300131x (2012).

40 Golbraikh, A. & Tropsha, A. Beware of q(2)! *J Mol Graph Model* **20**, 269-276, doi:Doi 10.1016/S1093-3263(01)00123-1 (2002).

41 Braga A.C., O. P. Diagnostic analysis based on ROC curves: theory and applications in medicine. *International Journal of Health Care Quality Assurance* **16**, 191-198 (2003).

42 Kolb, P. & Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J Med Chem* **49**, 7384-7392, doi:Doi 10.1021/Jm060838i (2006).

43 Majeux, N., Scarsi, M., Apostolakis, J., Ehrhardt, C. & Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins-Structure Function and Genetics* **37**, 88-105, doi:Doi 10.1002/(Sici)1097-0134(19991001)37:1<88::Aid-Prot9>3.0.Co;2-O (1999).

44 Majeux, N., Scarsi, M. & Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins-Structure Function and Genetics* **42**, 256-268, doi:Doi 10.1002/1097-0134(20010201)42:2<256::Aid-Prot130>3.0.Co;2-4 (2001).

45 *RDKit, Open-Source Cheminformatics.*, < http://www.rdkit.org.> (

46 Trott, O. & Olson, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **31**, 455-461, doi:Doi 10.1002/Jcc.21334 (2010).

47 Ashburn, T. T. & Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews. Drug discovery* **3**, 673-683, doi:10.1038/nrd1468 (2004).

48 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *Journal of molecular biology* **215**, 403-410, doi:DOI 10.1006/jmbi.1990.9999 (1990).

49 Doolittle, R. F. *Of Urfs and Orfs: Primer on how to analyze derived animo acid sequences*. (1986).

50 Rost, B. Twilight zone of protein sequence alignments. *Protein engineering* **12**, 85-94, doi:DOI 10.1093/protein/12.2.85 (1999).

51 Chung, S. Y. & Subbiah, S. A structural explanation for the twilight zone of protein sequence homology. *Structure* **4**, 1123-1127, doi:Doi 10.1016/S0969-2126(96)00119-0 (1996).

52 Gerstein, M. & Levitt, M. Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci* **7**, 445-456 (1998).

53 Sali, A. & Blundell, T. L. Comparative Protein Modeling by Satisfaction of Spatial Restraints. *Journal of molecular biology* **234**, 779-815, doi:DOI 10.1006/jmbi.1993.1626 (1993).

54 Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 5**, Unit 5 6, doi:10.1002/0471250953.bi0506s15 (2006).

55 Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-Dependent X-Ray-Diffraction as a Probe of Protein Structural Dynamics. *Nature* **280**, 558-563, doi:Doi 10.1038/280558a0 (1979).

56 Schuler, B. Single-molecule FRET of protein structure and dynamics - a primer. *Journal of nanobiotechnology* **11 Suppl 1**, S2, doi:10.1186/1477-3155-11-S1-S2 (2013).

57 Kay, L. E. NMR studies of protein structure and dynamics. *Journal of magnetic resonance* **173**, 193-207, doi:10.1016/j.jmr.2004.11.021 (2005).

58 Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nat Struct Biol* **7**, 740-743, doi:10.1038/78963 (2000).

59 Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646-652, doi:Doi 10.1038/Nsb0902-646 (2002).

60 MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **102**, 3586-3616 (1998).

61 Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* **25**, 1400-1415, doi:Doi 10.1002/Jcc.20065 (2004).

62 Vanommeslaeghe, K. *et al.* CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem* **31**, 671-690, doi:Doi 10.1002/Jcc.21367 (2010).

63 Yesselman, J. D., Price, D. J., Knight, J. L. & Brooks, C. L., 3rd. MATCH: an atom-typing toolset for molecular mechanics force fields. *J Comput Chem* **33**, 189-202, doi:10.1002/jcc.21963 (2012).

64 Krivov, S. V. & Karplus, M. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *The journal of physical chemistry. B* **110**, 12689-12698, doi:10.1021/jp060039b (2006).

65 Krivov, S. V. & Karplus, M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14766-14770, doi:10.1073/pnas.0406234101 (2004).

2. *IN SILICO* IDENTIFICATION OF ARTD-1 SELECTIVE INHIBITORS

2.1 INTRODUCTION

2.1.1 ARTD FAMILY OF PROTEINS

Poly(ADP-ribose) polymerase (PARP) family members PARP-1 and PARP-2 are nuclear proteins belonging to the ADP-ribosyl transferase family of enzymes. Using nicotinamide adenine dinucleotide (NAD^+) as substrate, they synthesize nicotinamide along with a negatively-charged ADP-ribose homopolymer (containing up to 200 ADP-ribose units and generally branched), formed in covalent attachment to one or more nuclear receptor proteins [1-4].

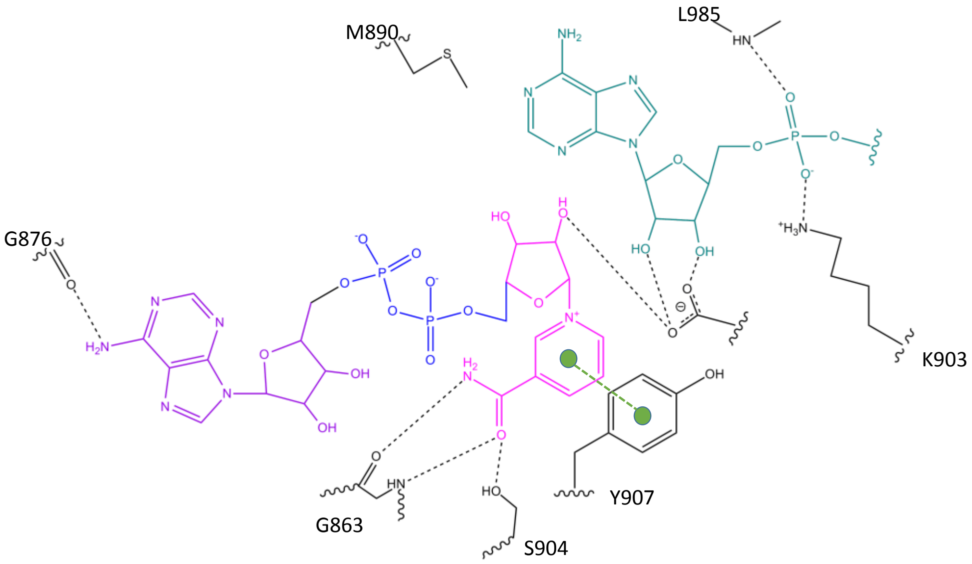


Figure 13. ARTD-1 binding interactions with NAD^+ donor and acceptor substrates adapted from [5]
Acceptor in cyan, donor nicotinamide in pink, donor phosphate in blue, donor adenine in purple

PARP-1 (116 kDa) was the first PARP family member to be identified in a 1963 publication referring to its activity [6], and was generally referred to as simply PARP up until the identification of further family members, including PARP-2 (62 kDa) in 1999 [7]. Today, 17 PARP proteins are known [8]. PARPs 1 and 2 synthesize poly(ADP-ribose) in response to DNA strand breaks [9,10] and thus participate in a signaling mechanism of a eukaryotic cell's recovery process. They have since been renamed ARTD-1 and ARTD-2 (ADP-ribosyltransferase diphtheria toxin-like) following their grouping into a larger family of structurally similar enzymes with catalytic functions leading to addition of a monomer or polymer of ADP-ribose to the recipient protein [11]. This new nomenclature will be applied hereon in.

ARTD-1 comprises three domains [12,13]. The N-terminal domain includes zinc fingers binding DNA at sites of strand breaks [14-17]. This is neighbor to the “automodification” domain (AD), which can be the acceptor of another ARTD protein’s ADP-ribose polymer. Finally, a highly conserved catalytic domain having a characteristic β-α-loop-β-α motif that includes the binding site of the donor NAD^+ forms the C-terminus [18,19].

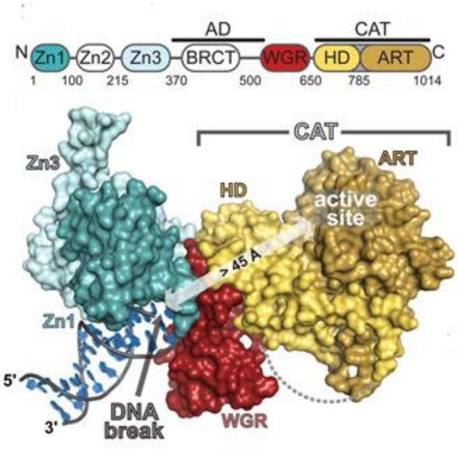


Figure 14. ARTD-1 domains from [17]

Following DNA binding, a rapid up-to 100-fold activation takes place resulting in poly(ADP-ribosyl)ation of receptor proteins, most of which are themselves ARTD proteins, but also include histones and polymerases. Automodification results in inactivation of the ARTD recipient. Addition of the poly(ADP-ribosyl) polymer to receptor proteins triggers chromatin structure relaxation and results in recruitment of the base excision repair (BER) proteins, including XRCC1 and DNA ligase III. Overactivation of ARTD-1, however, may result in cellular depletion of NAD^+ leading to necrosis [20]. Hence, ARTD-1 has two seemingly contradictory roles as a function of DNA damage extent: assisting in the recruitment of DNA repair proteins or inducing cell death by necrosis.

The above suggests two motivations to inhibit ARTD-1 enzymatic activity. The first is to limit necrosis-inducing overactivation of ARTD-1 in otherwise healthy cells during blood reperfusion following ischemia (i.e. subsequent to a stroke or heart attack), an approach that has been validated in mice models [21]. The second is to inhibit the DNA-repair function of ARTD-1 in diseased cells in order to support programmed cell death. Indeed, it is interesting to consider the role of ARTD-1 in cancer cells under two conditions. In general, alkylating chemotherapeutics and radiation induce the formation of DNA strand breaks [22]. If sufficient breaks accumulate in a diseased cell, desired cell death results. However, the effects of chemo- and radiation therapy are partially countered by ARTD's DNA repair activity. Inhibiting ARTD-1 in combination with chemo- and radiation therapy has been reported with some success, and knock-out mice deficient in either ARTD-1 or ARTD-2 are hypersensitive to alkylating agents and gamma-irradiation, both of which induce single-strand breaks [23].

The second condition considered desirable to target with ARTD-1 inhibitors are those involving cancer cells having defective DNA double strand break repair mechanisms. In presence of ARTD-1 inhibitors, DNA single strand breaks fail to be repaired and, during replication, become double strand breaks repaired by ARTD-independent mechanisms. BRCA-1 and BRCA-2 deficient tumor cells lack a functional double strand break repair pathway, such that ARTD-1 inhibitors have potential as single agent therapeutics. Healthy organisms treated with ARTD-1 inhibitors remain viable and ARTD-1 deficient mice were found to be free of abnormal rates of spontaneous tumor development [24].

2.1.2 ARTD-1 AND -2: SIMILAR PROTEINS, DIFFERING ROLES?

An estimated near 90% of nuclear ARTD activity is accounted for by ARTD-1 and the remaining 10% by ARTD-2. ARTD-2 shares many of the same receptor proteins as ARTD-1. However, different DNA structures may be recognised; indeed, this is unsurprising given that ARTD-2 lacks the traditional DNA binding domains of ARTD-1 (Fig. 16).

According to Schreiber [25], ARTD-2 binds DNA single strand breaks less efficiently than does ARTD-1, instead recognising gaps and flap structures. Though both ARTD-1 and ARTD-2 deficient phenotypes are hypersensitive to ionising radiation and alkylating agents, ARTD-2 deficient phenotypes appear less sensitive to low-dose radiation. In ARTD-1 deficient phenotypes, XRCC1 is recruited to DNA single strand breaks, whereas this recruitment is absent in ARTD-2 deficient phenotypes. Defects in spermatogenesis, adipogenesis and T cell development are seen in ARTD-2 but not ARTD-1 knockout mice, suggesting tissue-specific requirements for ARTD-2 [26,27]. ARTD-2, but not ARTD-1, is an important mediator of T-cell survival during thymopoiesis by preventing the activation of DNA-damage-dependant apoptotic responses [28]. Moreover, as discussed by Pellicciari, “the cell executioner role” cannot be taken over by ARTD-2 in ARTD-1 deficient animals [29]. The differing roles of ARTD-1 and ARTD-2 continue to be elucidated. That these two ARTDs may act at distinct steps and/or with distinct partners and/or in distinct subcellular localisations defines a motivation to develop inhibitors that are highly specific to each of these isoforms.

2.1.3 OVERVIEW OF KNOWN ARTD-1 AND ARTD-2 INHIBITORS

To date, most ARTD inhibitors mimic of the nicotinamide moiety, forming hydrogen bonds with Ser904/457 (ARTD-1/ARTD-2) and Gly863/429, and π-stacking interactions with Tyr907/473. Given that the mode of NAD^+ binding is highly similar in ARTD-1 and ARTD-2, inhibitors relying essentially on interactions with the nicotinamide binding site are not selective.

The few inhibitors reported to be selective are presented in Fig. 15. Of these, only one is accompanied by a co-crystal structure allowing confirmation of the nature of protein-ligand interactions including those that provide selectivity. Nonetheless, all structures suggest selective interactions outside of the nicotinamide binding site as discussed hereafter.

2.1.4 PROJECT OBJECTIVES

The objective of this work was to identify fragments and ligands binding to ARTD-1 with high potency (characterised by an IC_{50} or K_i of ≤ 10 µM) and forming interactions therewith not available in ARTD-2, such that these inhibitors are selective for ARTD-1. Selectivity is defined as a potency on ARTD-1 that is at least 10 fold greater than on ARTD-2. Such inhibitors have potential for use as tool compounds and possibly as therapeutic inhibitors. This work further more generally addresses the question of whether *in silico* methods can be used to identify new modes of selective interactions within a target binding site.

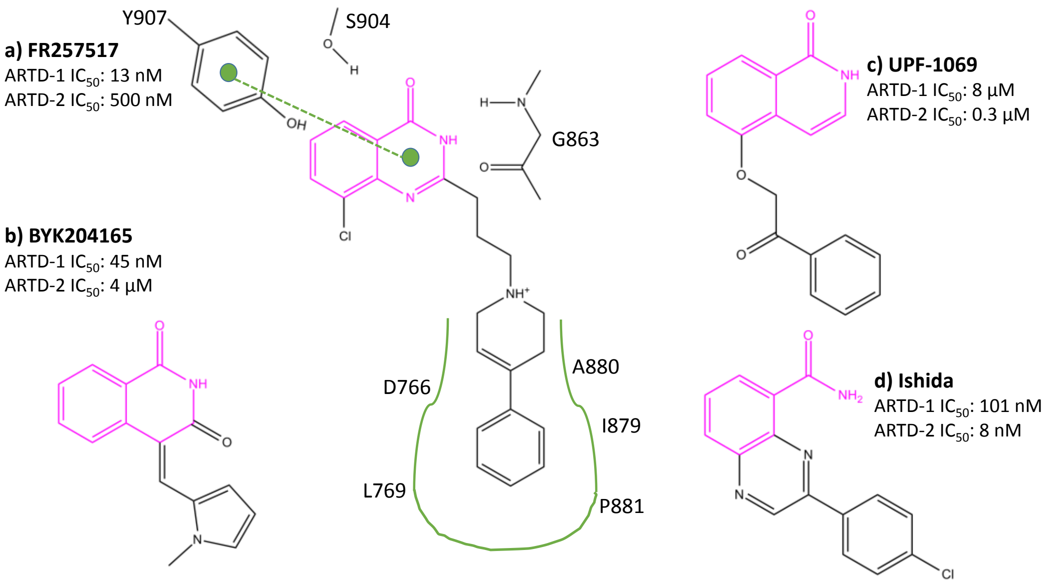
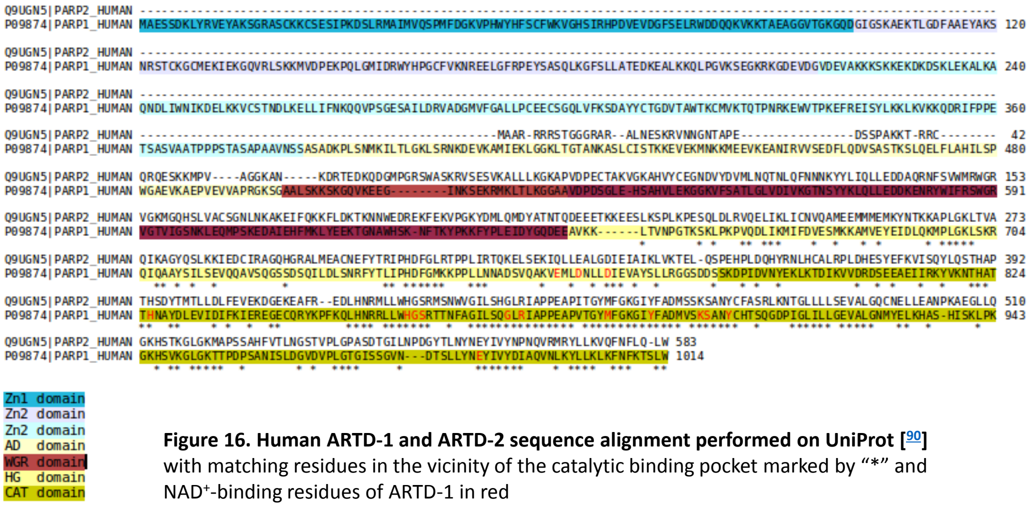


Figure 15. ARTD-1 and ARTD-2 selective inhibitors, with example of binding modes for compound a) having co-crystal structure with ARTD-1: ARTD-1 selective inhibitors a)[30] and b)[31], ARTD-2 selective inhibitors c) [32] and d) [33]. Nicotinamide binding moieties are in pink.

2.2 ARTD-1/-2 BINDING SITE COMPARISON

2.2.1 SEQUENCE COMPARISON AND STRUCTURAL ALIGNMENT

ARTD-1 and -2 catalytic domains exhibit high (64%) similarity. The catalytic pocket may be divided into sub-sites (Fig. 13). The acceptor site is occupied by the terminal ADP moiety of the poly(ADP-ribose) chain. The donor site is occupied by NAD$^+$, and may be further subdivided according to nicotinamide (NI), phosphate (PH) and adenine (AD) binding residues (Fig 17).



```
Q9UGN5|PARP2_HUMAN ------------------------------------------------------------------------------------------------------------------------
P09874|PARP1_HUMAN MAESSDKLYRVEYAKSGRASCKKCSESIPKDSLRMAIMVQSPMFDGKVPHWYHFSCFWKVGHSIRHPDVEVDGFSELRWDDQQKVKKTAEAGGVTGKGQDGIGSKAEKTLGDFAAEYAKS 120

Q9UGN5|PARP2_HUMAN ------------------------------------------------------------------------------------------------------------------------
P09874|PARP1_HUMAN NRSTCKGCMEKIEKGQVRLSKKMVDPEKPQLGMIDRWYHPGCFVKNREELGFRPEYSASQLKGFSLLATEDKEALKKQLPGVKSEGKRKGDEVDGVDEVAKKKSKKEKDKDSKLEKALKA 240

Q9UGN5|PARP2_HUMAN ------------------------------------------------------------------------------------------------------------------------
P09874|PARP1_HUMAN QNDLIWNIKDELKKVCSTNDLKELLIFNKQQVPSGESAILDRVADGMVFGALLPCEECSGQLVFKSDAYYCTGDVTAWTKCMVKTQTPNRKEWVTPKEFREISYLKKLKVKKQDRIFPPE 360

Q9UGN5|PARP2_HUMAN --------------------------------------------------MAAR-RRRSTGGGRAR--ALNESKRVNNGNTAPE----------------DSSPAKKT-RRC--------  42
P09874|PARP1_HUMAN TSASVAATPPPSTASAPAAVNSSASADKPLSNMKILTLGKLSRNKDEVKAMIEKLGGKLTGTANKASLCISTKKEVEKMNKKMEEVKEANIRVVSEDFLQDVSASTKSLQELFLAHILSP 480

Q9UGN5|PARP2_HUMAN QRQESKKMPV----AGGKAN-----KDRTEDKQDGMPGRSWASKRVSESVKALLLKGKAPVDPECTAKVGKAHVYCEGNDVYDVMLNQTNLQFNNNKYYLIQLLEDDAQRNFSVWMRWGR 153
P09874|PARP1_HUMAN WGAEVKAEPVEVVAPRGKSGAALSKKSKGQVKEEG--------INKSEKRMKLTLKGGAAVDPDSGLE-HSAHVLEKGGKVFSATLGLVDIVKGTNSYYKLQLLEDDKENRYWIFRSWGR 591

Q9UGN5|PARP2_HUMAN VGKMGQHSLVACSGNLNKAKEIFQKKFLDKTKNNWEDREKFEKVPGKYDMLQMDYATNTQDEEETKKEESLKSPLKPESQLDLRVQELIKLICNVQAMEEMMMEMKYNTKKAPLGKLTVA 273
P09874|PARP1_HUMAN VGTVIGSNKLEQMPSKEDAIEHFMKLYEEKTGNAWHSK-NFTKYPKKFYPLEIDYGQDEEAVKK------LTVNPGTKSKLPKPVQDLIKMIFDVESMKKAMVEYEIDLQKMPLGKLSKR 704
                                                                                         *       * *   ** ***    *  *   * *      * *****

Q9UGN5|PARP2_HUMAN QIKAGYQSLKKIEDCIRAGQHGRALMEACNEFYTRIPHDFGLRTPPLIRTQKELSEKIQLLEALGDIEIAIKLVKTEL-QSPEHPLDQHYRNLHCALRPLDHESYEFKVISQYLQSTHAP 392
P09874|PARP1_HUMAN QIQAAYSILSEVQQAVSQGSSDSQILDLSNRFYTLIPHDFGMKKPPLLNNADSVQAKVEMLDNLLDIEVAYSLLRGGSDDSSKDPIDVNYEKLKTDIKVVDRDSEEAEIIRKYVKNTHAT 824
                   ** * *  *         *             ** ******   ***         *   *  * *** *  *       *   * *  *          *  * *   *  *   ***

Q9UGN5|PARP2_HUMAN THSDYTMTLLDLFEVEKDGEKEAFR--EDLHNRMLLWHGSRMSNWVGILSHGLRIAPPEAPITGYMFGKGIYFADMSSKSANYCFASRLKNTGLLLLSEVALGQCNELLEANPKAEGLLQ 510
P09874|PARP1_HUMAN THNAYDLEVIDIFKIEREGECQRYKPFKQLHNRRLLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMVSKSANYCHTSQGDPIGLILLGEVALGNMYELKHAS-HISKLPK 943
                   **  **    * *  *  **         **** *******  *  *************** ************** *******  *    ****** *****   **  *      *

Q9UGN5|PARP2_HUMAN GKHSTKGLGKMAPSSAHFVTLNGSTVPLGPASDTGILNPDGYTLNYNEYIVYNPNQVRMRYLLKVQFNFLQ-LW 583
P09874|PARP1_HUMAN GKHSVKGLGKTTPDPSANISLDGVDVPLGTGISSGVN---DTSLLYNEYIVYDIAQVNLKYLLKLKFNFKTSLW 1014
                     * ** ****** *       * *  ****     *          *******    * ****  ***   **
```

```
Zn1 domain
Zn2 domain
Zn2 domain
AD  domain
WGR domain
HG  domain
CAT domain
```

Figure 16. Human ARTD-1 and ARTD-2 sequence alignment performed on UniProt [90] with matching residues in the vicinity of the catalytic binding pocket marked by “*” and NAD$^+$-binding residues of ARTD-1 in red

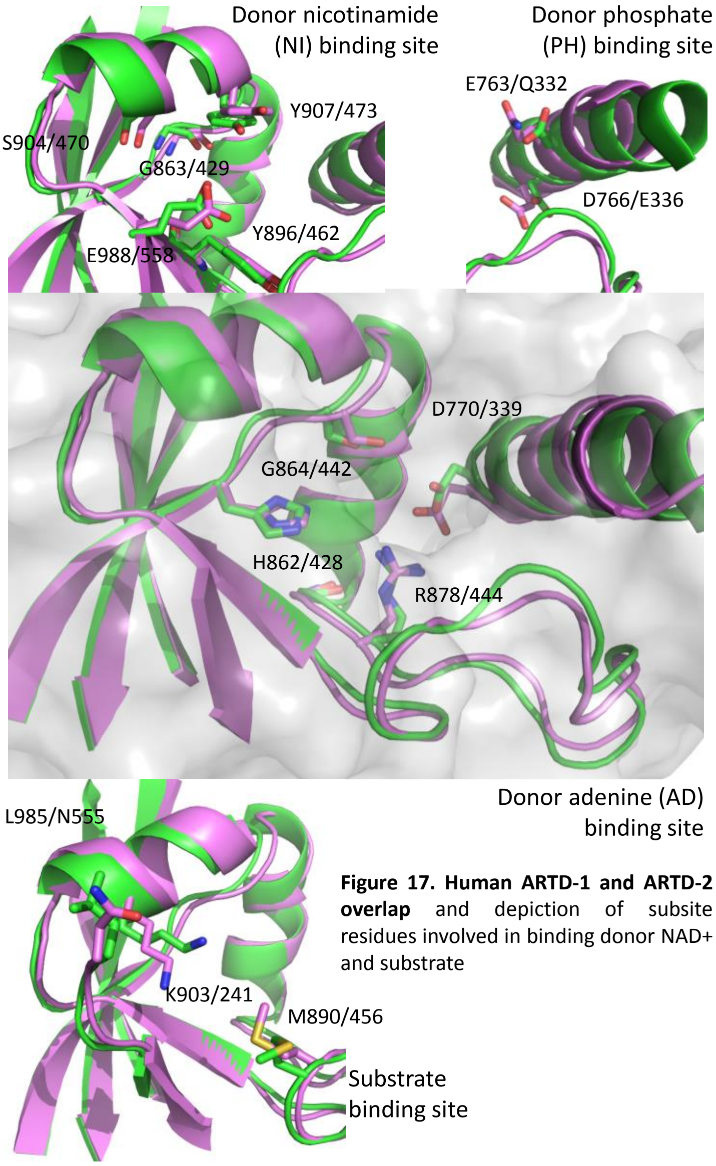


Figure 17. Human ARTD-1 and ARTD-2 overlap and depiction of subsite residues involved in binding donor NAD+ and substrate

To determine positional equivalence of binding site residues, human ARTD-1 and ARTD-2 crystal structures 3GJW and 3KJD, respectively, were overlaid using WITNOTP [34] giving an RMSD of 0.995 Å. Fig. 17 depicts residues involved in binding NAD$^+$ and the acceptor substrate. These residues are largely conserved, and account for a significant portion of the binding site. Correspondence between further residues along the binding sites of ARTD-1 and ARTD-2 was additionally determined, over a total of 46 residues per site.

2.2.2 RESIDUE ACCESSIBILITY TO BINDING SITE: A MOLECULAR DYNAMICS INVESTIGATION

Residues equivalently located within the ARTD-1 and ARTD-2 binding sites as determined by the above structural alignment were investigated by molecular dynamics for their accessibility to intramolecular binding as measured by their solvent accessible surface area (SASA).

Briefly, water molecules, ions and ligand were removed from PDB structures 3GJW (ARTD-1, resolution 2.30 Å) and 3KJD (ARTD-2, resolution 1.95 Å), and each was protonated such that aspartate and glutamate side-chains were negatively charged, arginine and lysine side-chains positively charged, while histidines were maintained neutral. Missing ARTD-2 loop residues (Thr349 to Gln352) of 3KJD were generated using the backbone coordinates of the equivalent loop of overlaid ARTD-1; side-chain atom coordinates were built using the autopsf function of VMD [35] and loop atoms minimised with CHARMM [36]. The obtained proteins were each immersed in a rectangular box of pre-equilibrated (TIP3P) water molecules having a minimal distance of 12 Å between the boundary and any protein atom. Sodium and chloride ions were added to neutralize the systems and to reach an ion concentration of 150 mM. Parameters for the system were from the CHARMM27 forcefield. Following minimisation, 0.5 ns of NVT and subsequent 0.5 ns of NPT equilibration during which the protein heavy atoms and protein Cα atoms were, respectively, positionally-restrained, molecular dynamics trajectories totalling 50 and 100 ns for ARTD-2 and ARTD-1, respectively, were collected in the NPT ensemble with a 2 fs time step at constant temperature (310 K, using the Langevin thermostat with a coupling coefficient of 1 ps^{-1} applied to heavy atoms) and constant pressure (using the modified Nosé-Hoover method with Langevin dynamics to control barostat fluctuations, with target pressure: 1.01325 bar; piston period: 200 fs; damping time scale: 100 fs) as implemented in NAMD [37-39]. The van der Waals and short-range electrostatic interactions were determine up to a cut-off of 10 Å. Periodic boundary conditions were applied to avoid finite size effects, and long-range interactions were evaluated by particle-mesh Ewald (PME) summation.

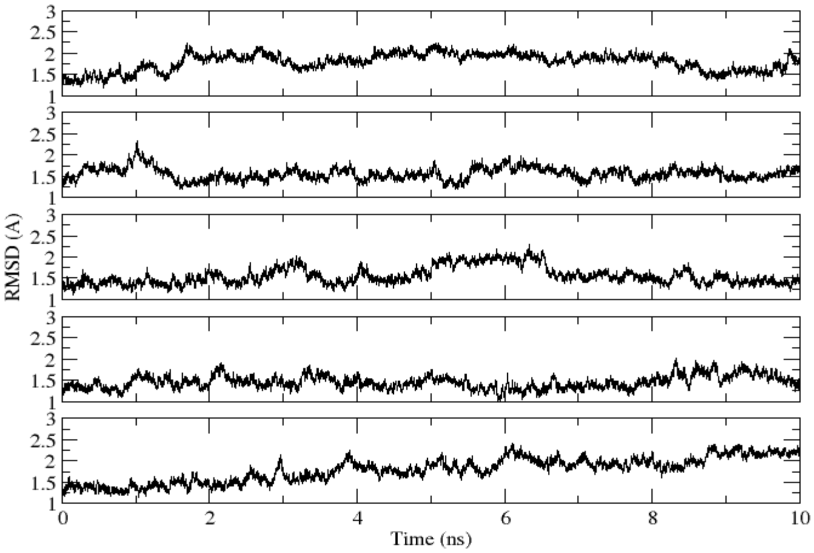


Figure 18. RMSD evolution over trajectories of ARTD-2: each panel represents a trajectory commenced from the same starting structure but with different seeds for the initial distribution of velocities.

Temporal protein conformation stability relative to the starting structure was evaluated based on the RMSD evolution of all Cα atoms as calculated with WORDOM [40], (Fig. 18). To determine the solvent accessible surface area (Fig. 19) of each residue side-chain along the binding sites of ARTD-1 and ARTD-2, the WORDOM [41] molecular surface (“ASURF”) calculation was run relative to side-chains atoms as of C_β using the Gepol numerical algorithm [42] with van der Waals radii from Rose [43] . The radius of the solvent sphere used to trace the surface was 1.4 Å and “NDIV” was set to 5.

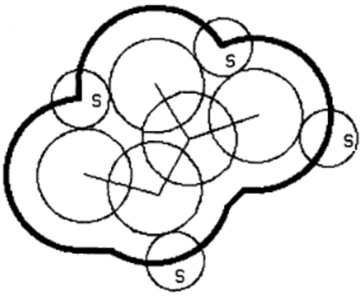


Figure 19. Solvent Accessible Surface Area from [42], wherein spheres indicated with "S" model the solvent and remaining spheres represent protein atoms; the thick line corresponds to the SASA (also referred to as ASURF)

Of the 46 investigated residues per binding site, those presenting physico-chemically different side-chains at equivalent positions in ARTD-1 and ARTD-2 where at least one of these side-chains is solvent accessible, as well as conserved residues with marked differences in solvent accessibility, were examined for their potential to interact with a ligand to form selective interactions. A set of these are shown in Fig. 21.

Interactions with Asp766/Glu355 have previously been exploited in the search for selective ligands [33,44]. Given the proximity of Asn767 to the nicotinamide binding site and the significant difference in possible interactions relative to its alanine counterpart in ARTD-2, this residue was selected for investigation of potential to form ARTD-1 selective hydrogen bonds with a ligand. It is of note that although ARTD-1 Asn767 is mentioned as an interaction partner of the quinazoline ligand of PDB 1UK0 [30,45], both structures of the unit cell of this PDB suggest that any such interaction is not direct, though possibly water-mediated.

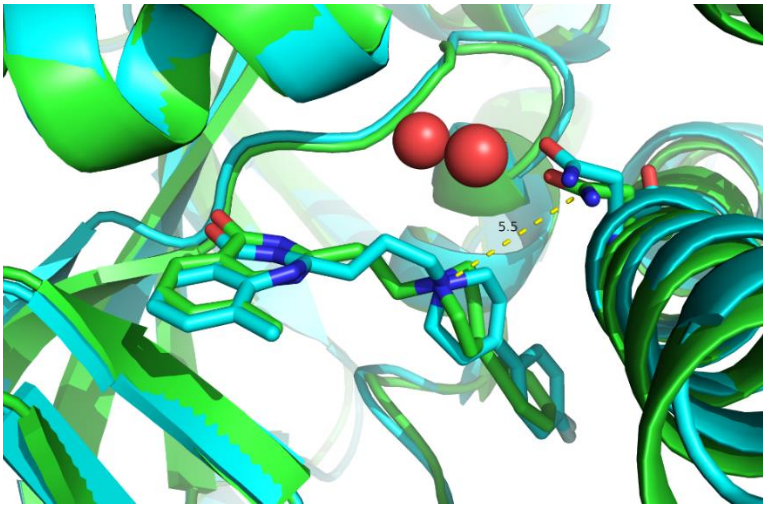


Figure 20. Backbone Cα overlaid structures of PDB 1UK0 unit cell with ligand and Asn767 in stick representation, and crystal water oxygens near to Asn767 in spherical representation; dashed lines show the measure of indicated distance in Å

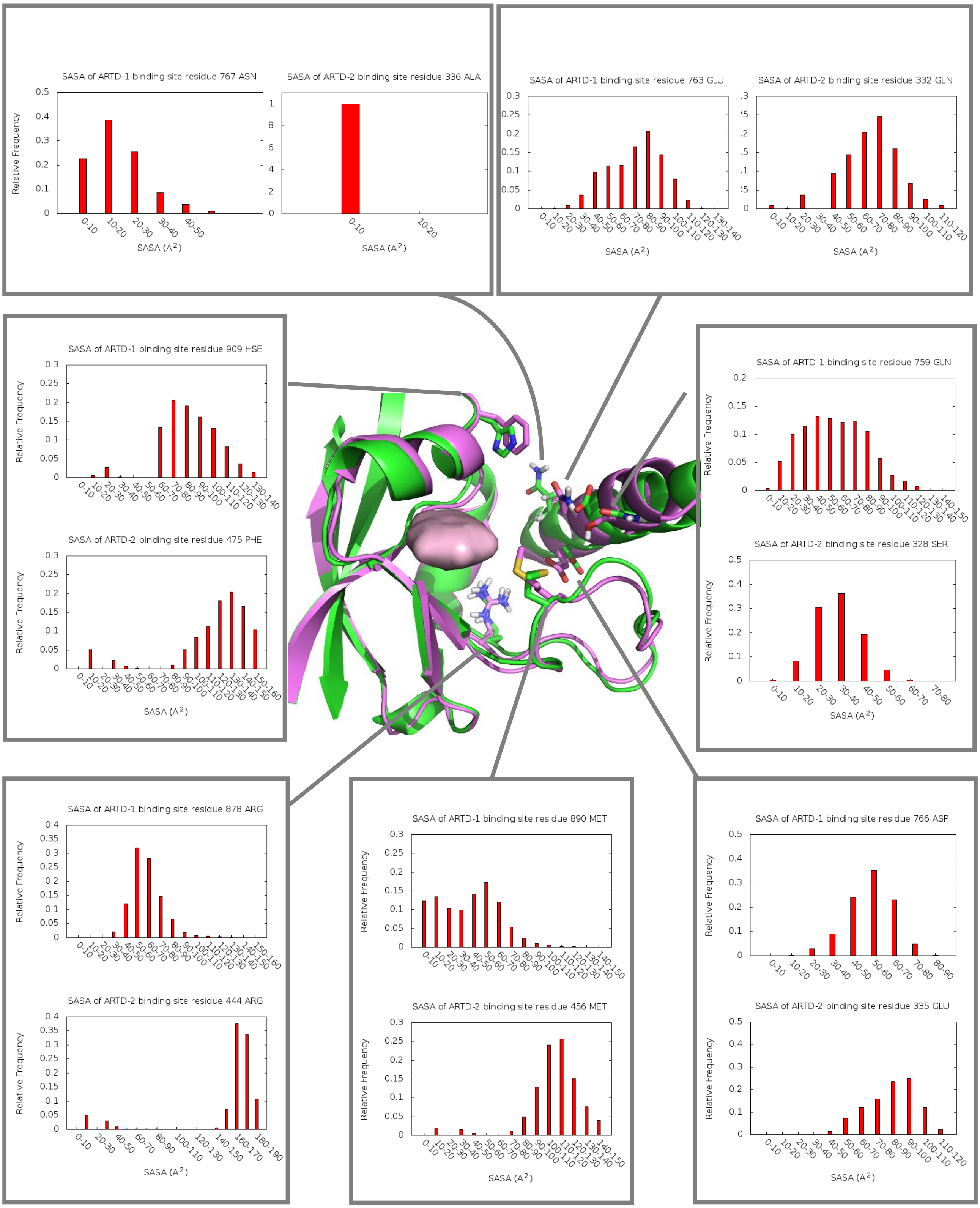


Figure 21. MD-based SASA distribution of residues providing potential basis for selective interactions with ARTD-1 (green) and ARTD-2 (magenta); the nicotinamide binding site is presented as a pink surface

2.3 ARTD-1 PROTEIN CONFORMATIONS FOR DOCKING

2.3.1 CUT-BASED FREE ENERGY PROFILE

Following selection of Asn767 as target residue of the ARTD-1 binding site, the ARTD-1 MD trajectory was further analysed to identify alternative, energetically-favourable binding site conformations in which Asn767 remains accessible, as evaluated based on its side-chain SASA. For this, a one-dimensional cut-based free energy profile (cFEP) with a normalised partition function as progress coordinate was prepared [46]. The input for the cFEP calculation is a network of nodes containing trajectory frames, joined by edges whose weights are proportional to the transitions between them. Trajectory frames were first clustered ("binned") based on RMSD using the leader algorithm in WORDOM relative to all non-symmetric heavy atoms of the binding site lying within 7 Å of Asn767. Node partitioning into two groups according to the minimum-cut procedure gives the partition function of the cutting surface separating them, from which the free energy barrier is derived. By iterative determination of the minimum cuts between all pairs of nodes, all barriers and basins are isolated. The RMSD cut-off value of 0.85 Å as per Fig. 22 was selected to ensure consistency with a diffusive regime as per [47]; use of this cut-off meets the requirement that F be shifted by 0.35(ln(√2)) consequent to the use of every second frame of the trajectory (a doubling of dt).

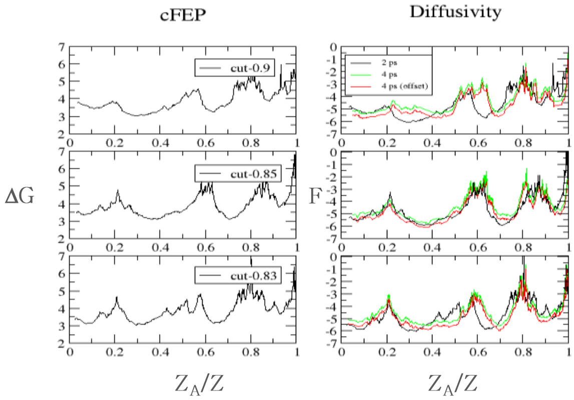


Figure 22. cut-based Free Energy Profile (cFEP) and diffusivity evaluation with different clustering cut-off ("cut") values

2.3.2 BINDING SITE RESIDUE DIHEDRAL ANGLES AND SASA WITHIN BASINS

Firstly, the SASA of the Asn767 side chain was determined over the cFEP in order to define basins in which its values were desirably high. Secondly, the dihedral angle values of side chains along the binding site were mapped along the cFEP and frames were selected to fall within populated values of each of these dihedrals with the respective free energy basins. A selection of main results from this analysis is given in Fig. 23.

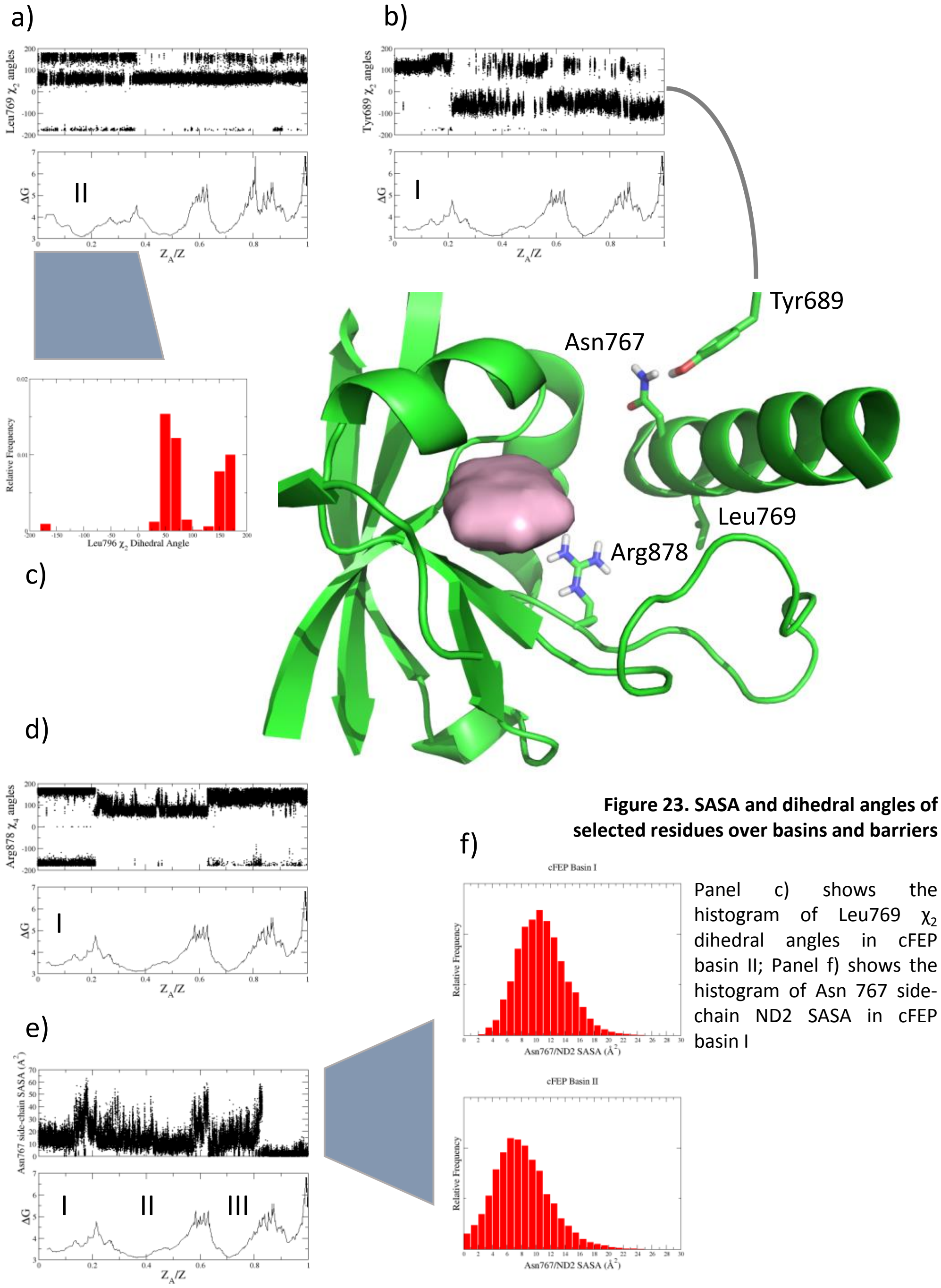


Figure 23. SASA and dihedral angles of selected residues over basins and barriers

Panel c) shows the histogram of Leu769 χ_2 dihedral angles in cFEP basin II; Panel f) shows the histogram of Asn 767 side-chain ND2 SASA in cFEP basin I

2.3.3 SUMMARY OF SELECTED PROTEIN CONFORMATIONS

Only three PDB structures of human ARTD-1 include Asn767 conformations providing non-zero side-chain atom SASA (PDB ID: 1UK0, 1UK1, 2RCW). The location of the closest MD-generated protein conformation (RMSD<1 Å) to each PDB conformation along the cFEP places all three in the third basin.

Accordingly, PDB structures 1UK0 and 2RCW were retained as representatives of basin III. Two additional MD-generated protein conformations were selected for docking, one from each of cFEP basins I and II with high Asn767 side-chain SASA and dihedral angles consistent with highly populated values in their respective basin.

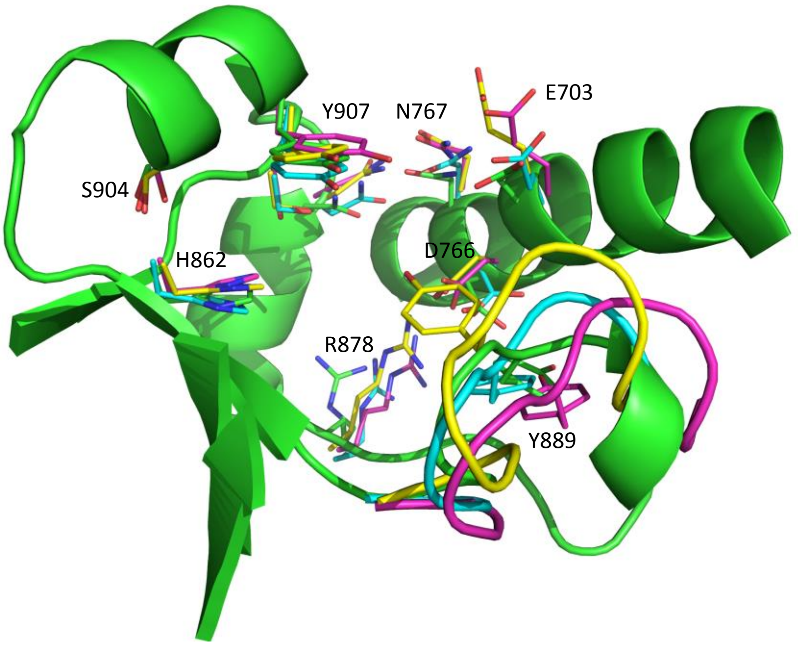


Figure 24. Selected protein conformations for docking: MD-generated frames were selected to represent basins I (pink) and II (yellow), while 2 PDB structures represent basin III (1UK0: green, 2RCW: blue)

2.4 ARTD-1 SCORING FUNCTION

2.4.1 TRAINING SET LIGANDS AND POSE GENERATION

A set of 32 inhibitors of ARTD-1 with reported K_i or IC_{50} values (ranging from 3 nM to 190 µM) and comprising a known nicotinamide binding moiety were used to build a scoring function to rank ligands in a high-throughput virtual screening campaign; this moiety requirement facilitated recognition of the probable binding mode.

Ligand poses were established in a single conformation of ARTD-1, namely that of PDB structure 1UK0. Use of a single structure removes variation in the subsequently-calculated electrostatic term due to differences in protein conformation outside of the binding site. The PDB 1UK0 protein structure was prepared by removing all water molecules, ions and ligand and subsequently protonating such that aspartate and glutamate side-chains were negatively charged, arginine and lysine side-chains positively charged, while histidines were maintained neutral. Hydrogen atom coordinates were minimised with CHARMM.

For ligands having a co-crystal structure with ARTD-1, poses were generated by backbone C_α atom overlay of the co-crystal protein with those of PDB structure 1UK0 using WITNOTP [34]. Remaining ligands were docked to the binding site of the protein conformation of PDB 1UK0 using VINA [48], and visually inspected for recapitulation of the expected binding mode at the nicotinamide binding pocket.

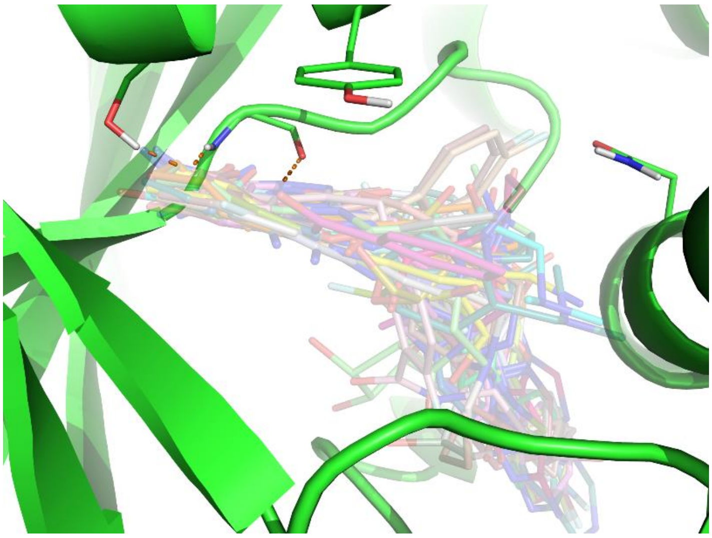


Figure 25. Binding modes of 32 ARTD-1 inhibitors used to develop a scoring function in the ARTD-1 binding site of PDB 1UK0
Hydrogen bonds to NI binding site are shown as dashed lines

2.4.2 EVALUATION OF TERMS

Protein atoms were parametrised with CHARMM27 parameters while the ligand was parametrised according to the CHARMM General Force Field [49] with MATCH [50]. Docked poses were minimised with CHARMM (by the iterative steepest descent method for 500 steps followed by the conjugate gradient method for 10'000 steps), allowing residues with atoms within 5 Å of the ligand to remain flexible along with the ligand during minimisation.

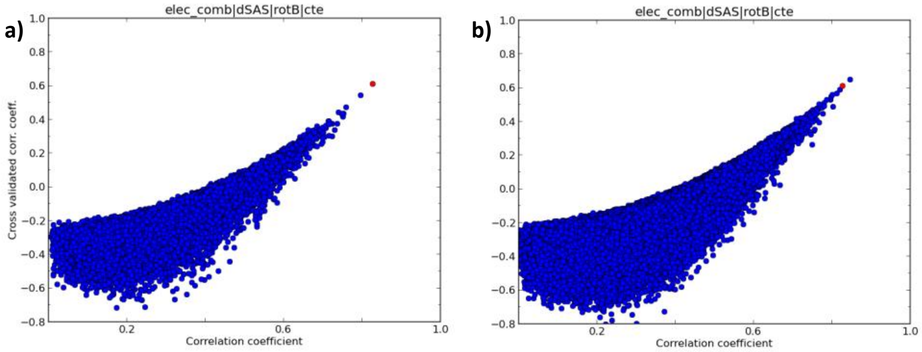
A model for binding affinity was then built by multivariate linear least squares regression to determine the coefficients based on a set of equations of the form:

$$\Delta G_{binding} \propto -RT\ln(\mathrm{IC}_{50,\exp}) = c_1 N_{rot} + c_2 \Delta SASA_{binding} + c_3 \Delta E_{vdW} + c_4 \Delta G_{electr} + c_5$$

The number of rotatable bonds (N_{rot}) was calculated by DAIM [51]. Change in SASA on binding (ΔSASA) was calculated using CHARMM [36] as the difference in SASA of the complex to the SASA of the free protein and ligand ($\Delta SASA = SASA_{complex} - SASA_{protein} - SASA_{ligand}$) using the "COOR SURF" command with a solvent diameter of 1.4 Å and atom radii from the CHARMM27 parameter file for the protein and CHARMM36 CGenFF parameter file for the ligand. The intermolecular van der Waals energy (ΔE_{vdW}) was calculated with CHARMM applying a switching function between 10 Å and the truncation distance of 12 Å. Electrostatic energy (ΔG_{electr}) was calculated according to the thermodynamic cycle of Fig. 9 with CHARMM, where ΔG_{electr} is a sum of two terms ($\Delta E_{Coulomb} + \Delta G_{solv}$): the Coulombic energy ($\Delta E_{Coulomb}$, calculated with fixed ε=2 and infinite evaluation cut-off) was evaluated relative to all atom pairs separated by at least 3 covalent bonds, and the free energy of desolvation on complex formation, calculated with the CHARMM PBEQ module and a focussing procedure with a final grid spacing of 0.3 Å. These latter two terms (ΔE_{vdW}, ΔG_{electr}) correspond to and were calculated consistently with those of the two-parameter model with rigorous treatment of continuum electrostatics (i.e. numerical solution of the Poisson equation by the finite difference method) reported by Huang [52].

2.4.3 SCORING FUNCTION: CORRELATION AND CROSS-VALIDATION CORRELATION

Multivariate linear least squares regression was performed including cross-validation by the leave-one-out procedure, applied in turn to (i.e. leaving out in turn) each of the input data sets. Models for all subsets of the set of X_i's were determined, with coefficients obtained for each model. The 3-parameter model given in c) of Fig. 26 had a cross-validation correlation coefficient (q^2) of 0.61.



c) $$\Delta \hat{G}_{binding} = 0.032(\Delta G_{electr}) + 0.014(\Delta SASA) + 0.72(N_{rot\_bonds}) - 1.51$$

Figure 26. Scoring function model c) and evaluation relative to a) 100'000 and b) 1 million models build with random values of ΔG within the ΔG range of the original set: the model of c) is represented by a red dot in panels a) and b) whereas random models are shown as blue dots

The model of Fig. 26c) was evaluated relative to sets of models (a) 100'000 and b) 1 million in Fig. 26) issued from datasets in which X_i values were maintained and ΔG values randomly assigned over the range of ΔG values of the original set. That the scoring function model remains largely superior with few exceptions even relative to 1 million such random models attests to its expected predictive power.

2.4.4 SCORE CUT-OFF DETERMINATION

The obtained 3-parameter model was used to plot experimental relative to calculated estimates of $\Delta G_{binding}$ as per Fig. 27. A stringent score cut-off of -9.0 is expected to provide minimal false-positives while capturing most true high affinity binders.

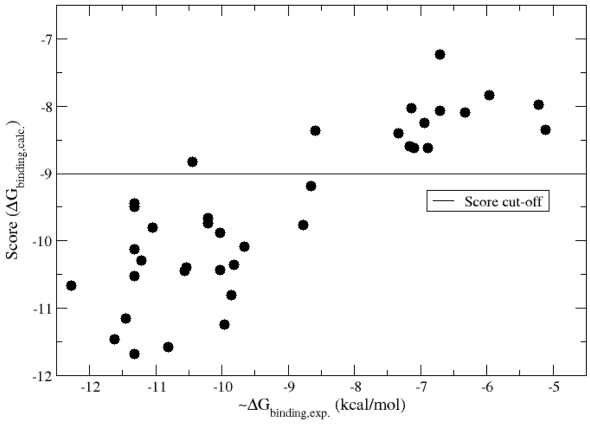


Figure 27. Experimental versus calculated estimates of $\Delta G_{binding}$ relative to ligands of the training set.

2.5 FRAGMENT-BASED LIBRARY DOCKING

2.5.1 MOTIVATION AND PROTOCOL

The objective of identifying ligands including a moiety binding to Asn767 with high binding affinity indicates the suitability of fragment-based docking to the pocket of this target residue, notably as implemented in the ALTA procedure [51] (Fig. 11).

2.5.2 FRAGMENT LIBRARY GENERATION

The ALTA procedure commences with a library of fragments defined as low molecular weight compounds with minimal rotatable bonds. The ZINC [53] lead-like library of near 4 million compounds (Mw< 350 g/mol, $N_{rot} \leq 7$) was accordingly decomposed with DAIM [51] and filtered for fragments containing at least one hydrogen bond donor or acceptor, thus obtaining a library of 108’000 fragments. Characteristics of this fragment library are given in Fig. 28.

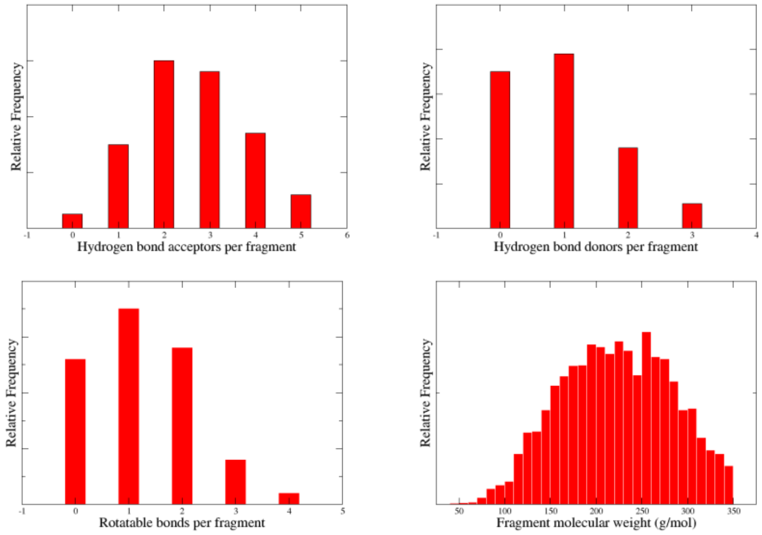


Figure 28. Fragment library characteristics

2.5.3 FRAGMENT DOCKING, RANKING AND SELECTION

Fragments were docked to each of three protein structures (2 MD-generated structures discussed previously and PDB structure 1UK0) using SEED [54,55] based on the polar docking vectors of Asn767 with 36 rotations around each docking vector and accurate calculation of SEED energy terms: van der Waals interactions and electrostatic energy with continuum solvation. A SEED total energy cut-off of -5 kcal/mol was applied to the top-ranked pose of each fragment; the resulting set of fragments were then ordered by their median rank across six terms: electrostatic difference, transfer energy, total energy, total energy efficiency (total energy/heavy atom count), van der Waals efficiency (van der Waals energy/heavy atom count) and electrostatic efficiency (electrostatic energy/heavy atom count) [56]. A total of 50 fragments were selected (a subset of which are given in Fig. 30) relative to which a substructure search of the ZINC all purchasable library of 19 million compounds was performed with RDKIT [57]. It is of note that the fragments selected by each of the three protein conformations were largely unique (Fig. 29).

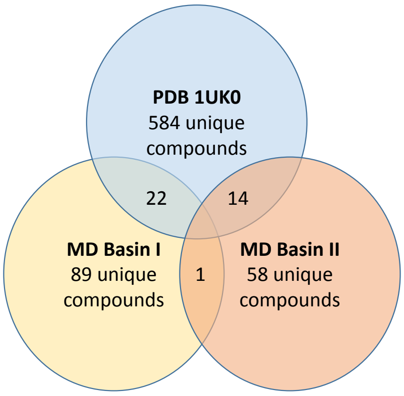


Figure 29. Uniqueness of fragments binding to Asn767 discovered by SEED

2.5.4 FRAGMENT-CONTAINING LIGAND LIBRARY DOCKING AND SELECTION

The substructure search lead to a library of near 800'000 compounds. Compounds of this library were protonated relative to a pH of 7 using obabel [58]. These were docked to the protein conformations relative to which the respective fragments were selected using VINA [48] to generate 20 poses per compound with filtering for a hydrogen bond to Asn767 and at least one further hydrogen around the nicotinamide binding site [59]. Issued poses were then filtered for reproduction of the anchor fragment binding mode by the corresponding compound moiety using RDKIT (RMSD of fragment to compound moiety of less than or equal to 3 Å) prior to minimisation with CHARMM (performed as described 2.4.2) and scoring (as per 2.4.3).

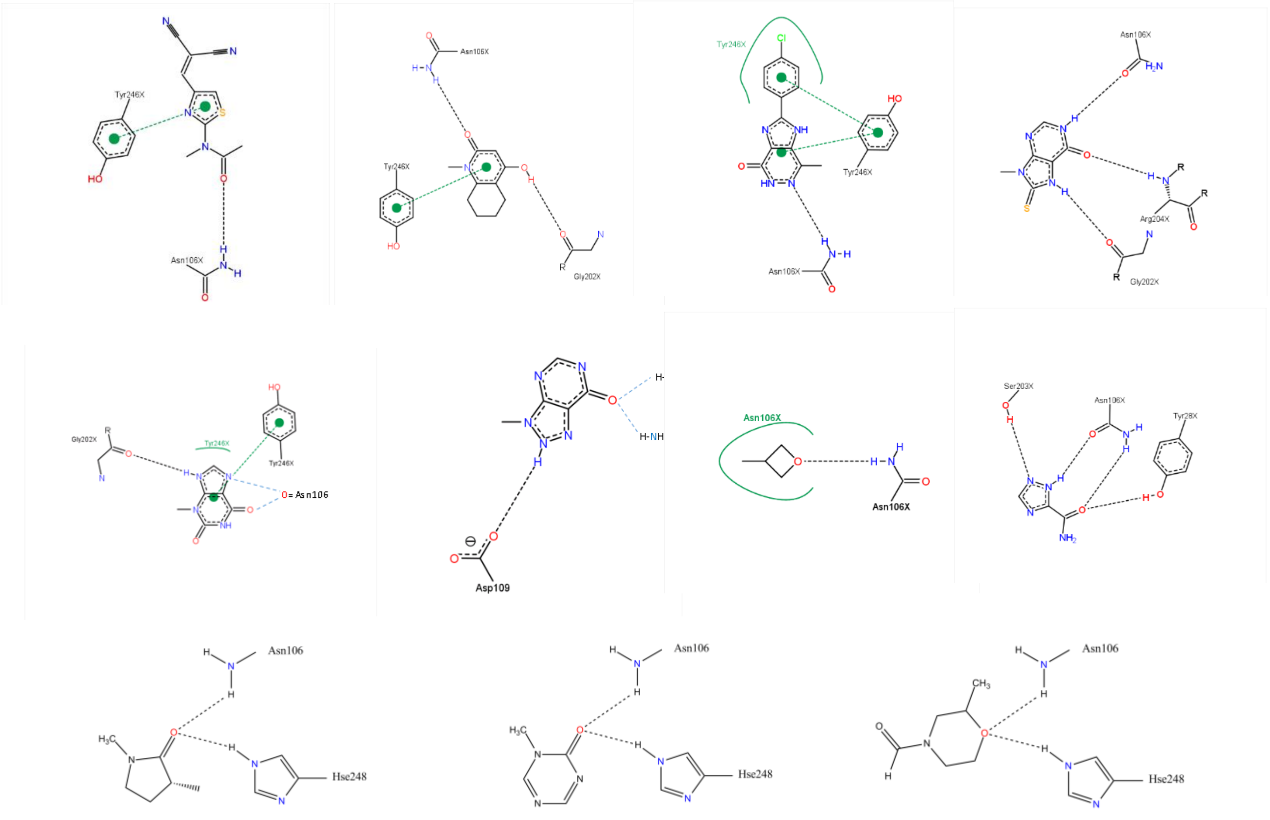


Figure 30. Selection of fragments binding to Asn767 discovered by SEED
With residue numbering shifted by -661

2.5.5 SUMMARY OF SELECTED LIGANDS

A total of 350 unique compounds passed the scoring cut-off. Two example compounds with their overlaid anchor fragment poses from SEED are given in Fig. 31. A subset of these compounds is currently under evaluation of binding affinity to ARTD-1 with an external service provider. Potent compounds will thereafter be evaluated for selectivity on ARTD-2.

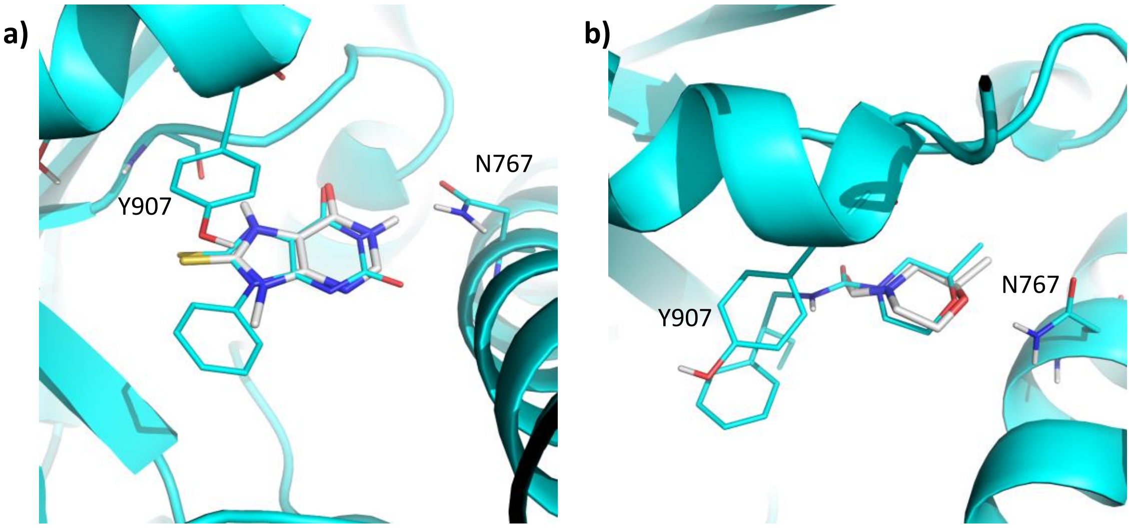


Figure 31. Putative selective ligands of ARTD-1 identified with the ALTA procedure;
a) shows a ligand (cyan) selected relative to the protein conformation of PDB 1UK0 overlapped with its anchor identified by fragment docking, b) shows another ligand docked into MD-generated protein conformation from Basin II.

2.6 NICOTINAMIDE-MIMIC LIBRARY DOCKING

2.6.1 MOTIVATION AND PROTOCOL

Fragment binding to the nicotinamide binding site contribute highly to binding affinity. Indeed, a number of low molecular weight ligands binding uniquely at this site have low micromolar binding affinities [19,60]. As any ligand providing selective binding to Asn767 of ARTD-1 should also be potent, a library of compounds comprising such moieties was assembled. These moieties also provide the advantage of characteristic binding modes allowing reasoned selection of putative binding poses following docking.

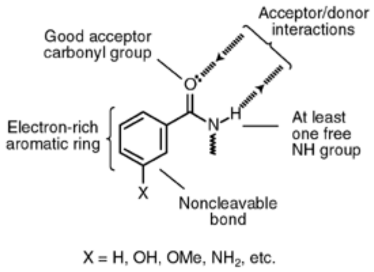


Figure 32. Structural requirements for potent ARTD-1 inhibitors from [61]
showing interactions at the nicotinamide binding site

2.6.2 NICOTINAMIDE-MIMIC LIBRARY GENERATION

Known nicotinamide binding site moieties were assemble from the literature and are summarised in Fig. 33 [61-65]. These moieties were used to perform a substructure search of the eMolecules database [66] with the requirement that the characteristic NH group of Fig. 33 not be replaced by a heavy atom in the selected ligands. This resulted in a library of near 30’000 compounds. Compounds of the library were protonated relative to a pH of 7 using obabel [58].

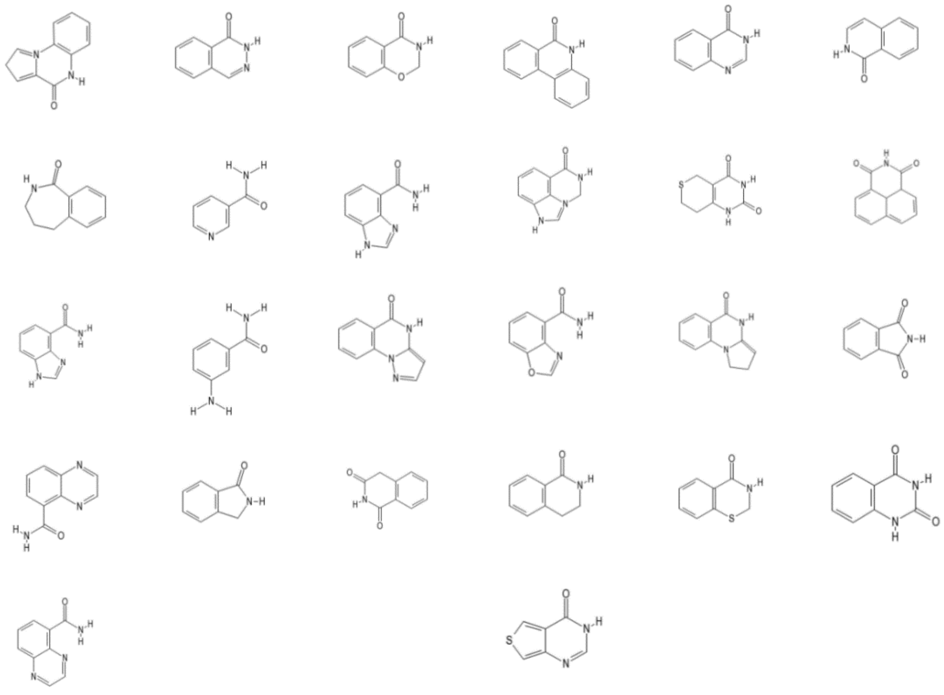


Figure 33. Fragments reported to bind the nicotinamide binding site used for substructure searching

2.6.3 LIGAND DOCKING AND SELECTION

Ligands were docked to the four protein conformations of 2.3.3 using VINA [48] to generate 20 poses per compound, with filtering for a hydrogen bond to Asn767 and at least one further hydrogen bond to the nicotinamide binding site. Filtered poses were then minimised with CHARMM (as per 2.4.2), with scoring and application of the scoring cut-off resulting in 141 unique compounds. These were then clustered based on the following classes of moieties interacting with Asn767 with examples given in Fig. 34: a) conjugated amides including benzamides (14 compounds), b) pyridines (4 compounds), c) methoxyphenyls (24 compounds), d) 5-membered heterocycles comprising N-0 ring atoms oriented towards Asn767 (16 compounds), e) nitriles (3 compounds), f) non-conjugated amides (4 compounds), g) sulfones and sulfonamides (7 compounds), h) carboxylates (5 compounds), i) non-conjugated ethers including cyclic ethers (6 compounds), j) conjugated ethers including furans and dioxolanes (9 compounds), k) 5-membered heterocycles containing N-N oriented towards Asn767, including triazoles (4 compounds), as well as N-S 5-membered heterocycles, pyrroles and anilines. Selected ligands from the different moiety classes are currently undergoing evaluation at an external service provider.

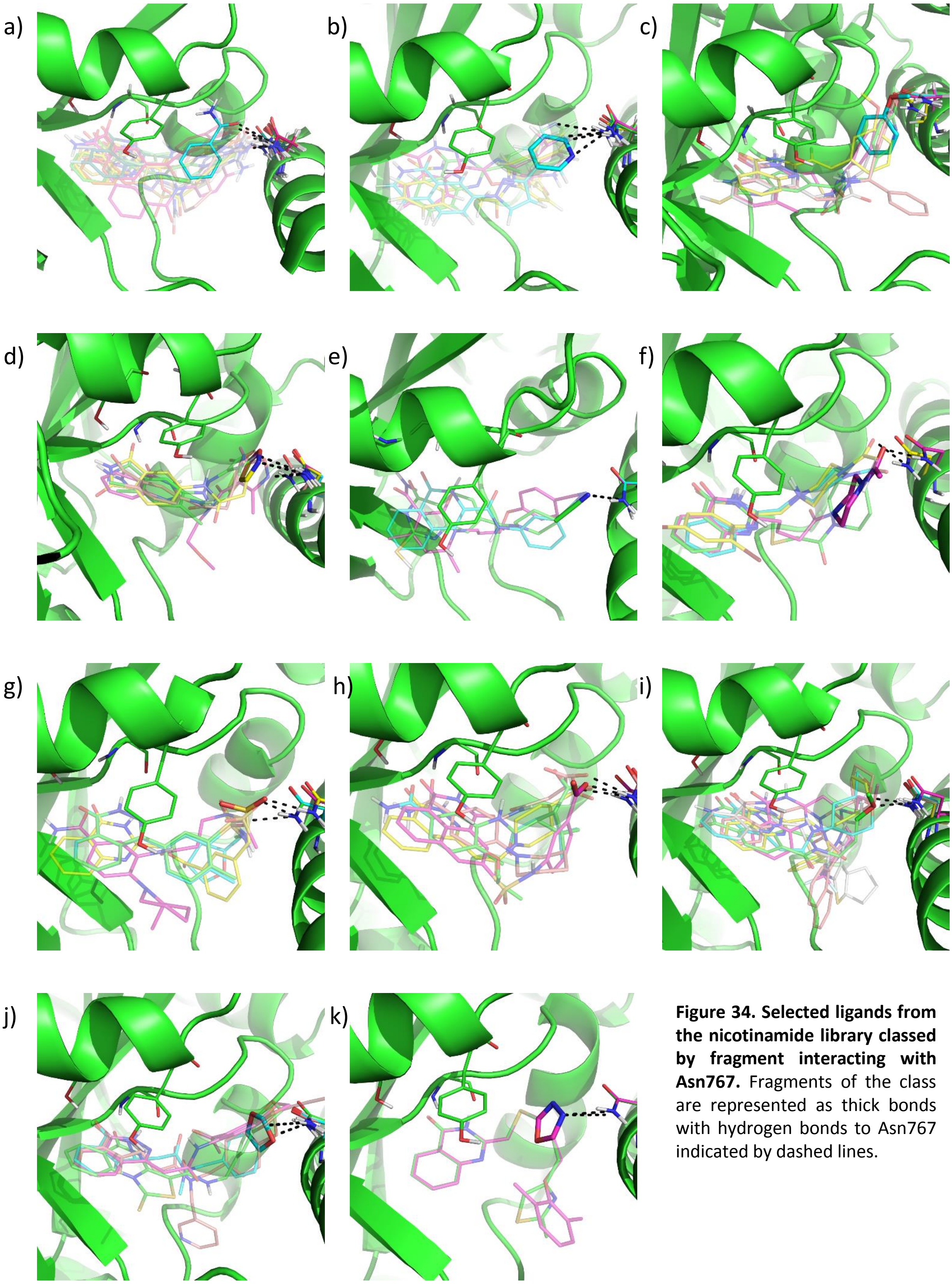


Figure 34. Selected ligands from the nicotinamide library classed by fragment interacting with Asn767. Fragments of the class are represented as thick bonds with hydrogen bonds to Asn767 indicated by dashed lines.

2.7 DISCUSSION

In silico methods may provide particular advantages over their purely experimental counterparts in the search for novel compounds forming interactions with particular residues or subsites of a target protein. Indeed, their atomic level tailoring allows rational selection of compounds prior to experimental validation. In particular, fragment-based approaches such as ALTA can be used to identify anchor fragments binding with high affinity to a target site.

In the present study, two approaches were used to identify fragments and compounds binding to a new residue identified a basis for potential selective interactions with ARTD-1 relative to ARTD-2, namely a fragment-based approach and a tailored library approach. Each one provided a series of moieties and molecules containing these moieties.

Molecular dynamics was used before docking to establish the accessibility of binding site residues as well as differential accessibility of equivalent residues between ARTD-1 and ARTD-2. This critical information is not yet available by other techniques, in which the dynamic aspects of protein binding sites are often observed only once inhibitors stabilising alternative protein conformations are at hand. As per McCammon's [67] statement that "[v]irtual screening results are influenced by the PDB structure chosen for screening because the fit of individual ligands in the pocket is affected by even the most minor structural changes", we found that small changes in protein conformation strongly influenced selected fragments and ligands passing a set cut-off. This underlines the importance of including a variety of protein structures in a high-throughput docking campaign.

2.8 CONCLUSIONS

The ligands proposed as putative high-affinity binders of ARTD-1 that are selective over ARTD-2 are currently undergoing experimental validation. The results will form the basis of follow-up studies, including crystal structure validation of *in silico* generated binding modes, offering insights into the use of molecular dynamics to identify binding-site accessible residues and novel protein conformations of relevance to high-throughput screening. Any such promising ligands will be considered for further development as potential tool or therapeutic compounds.

2.8 REFERENCES

1 Hayaishi, O. & Ueda, K. Poly(Adp-Ribose) and Adp-Ribosylation of Proteins. *Annual review of biochemistry* **46**, 95-116, doi:DOI 10.1146/annurev.bi.46.070177.000523 (1977).

2 Ueda, K., Ogata, N., Kawaichi, M., Inada, S. & Hayaishi, O. ADP-ribosylation reactions. *Current topics in cellular regulation* **21**, 175-187 (1982).

3 Mandel, P., Okazaki, H. & Niedergang, C. Poly(adenosine diphosphate ribose). *Progress in nucleic acid research and molecular biology* **27**, 1-51 (1982).

4 Ueda, K. & Hayaishi, O. ADP-ribosylation. *Annual review of biochemistry* **54**, 73-100, doi:10.1146/annurev.bi.54.070185.000445 (1985).

5 Jagtap, P. & Szabo, C. Poly(ADP-ribose) polymerase and the therapeutic effects of its inhibitors. *Nature Reviews Drug Discovery* **4**, 421-440, doi:Doi 10.1038/Nrd1718 (2005).

6 Chambon, P., Weill, J. D. & Mandel, P. Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme. *Biochemical and biophysical research communications* **11**, 39-43 (1963).

7 Ame, J. C. *et al.* PARP-2, A novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase. *The Journal of biological chemistry* **274**, 17860-17868 (1999).

8 Hassa, P. O. & Hottiger, M. O. The diverse biological roles of mammalian PARPS, a small but powerful family of poly-ADP-ribose polymerases. *Frontiers in bioscience : a journal and virtual library* **13**, 3046-3082 (2008).

9 Benjamin, R. C. & Gill, D. M. Poly(Adp-Ribose) Synthesis Invitro Programmed by Damaged DNA - a Comparison of DNA-Molecules Containing Different Types of Strand Breaks. *Journal of Biological Chemistry* **255**, 502-508 (1980).

10 Benjamin, R. C. & Gill, D. M. Adp-Ribosylation in Mammalian-Cell Ghosts - Dependence of Poly(Adp-Ribose) Synthesis on Strand Breakage in DNA. *Journal of Biological Chemistry* **255**, 493-501 (1980).

11 Hottiger, M. O., Hassa, P. O., Luscher, B., Schuler, H. & Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends in biochemical sciences* **35**, 208-219, doi:10.1016/j.tibs.2009.12.003 (2010).

12 Kurosaki, T. *et al.* Primary structure of human poly(ADP-ribose) synthetase as deduced from cDNA sequence. *The Journal of biological chemistry* **262**, 15990-15997 (1987).

13 Ame, J. C., Spenlehauer, C. & de Murcia, G. The PARP superfamily. *Bioessays* **26**, 882-893, doi:Doi 10.1002/Bies.20085 (2004).

14 D'Silva, I. *et al.* Relative affinities of poly(ADP-ribose) polymerase and DNA-dependent protein kinase for DNA strand interruptions. *Bba-Protein Struct M* **1430**, 119-126, doi:Doi 10.1016/S0167-4838(98)00278-7 (1999).

15 Pion, E. *et al.* Poly(ADP-ribose) polymerase-1 dimerizes at a 5 ' recessed DNA end in vitro: A fluorescence study. *Biochemistry-Us* **42**, 12409-12417, doi:Doi 10.1021/Bi0344432 (2003).

16 Lonskaya, I. *et al.* Regulation of poly(ADP-ribose) polymerase-1 by DNA structure-specific binding. *Journal of Biological Chemistry* **280**, 17076-17083, doi:DOI 10.1074/jbc.M413483200 (2005).

17 Langelier, M. F., Planck, J. L., Roy, S. & Pascal, J. M. Structural Basis for DNA Damage-Dependent Poly(ADP-ribosyl)ation by Human PARP-1. *Science* **336**, 728-732, doi:DOI 10.1126/science.1216338 (2012).

18 Rolli, V., O'Farrell, M., Menissier-de Murcia, J. & de Murcia, G. Random mutagenesis of the poly(ADP-ribose) polymerase catalytic domain reveals amino acids involved in polymer branching. *Biochemistry-Us* **36**, 12147-12154, doi:10.1021/bi971055p (1997).

19 Ruf, A., de Murcia, G. & Schulz, G. E. Inhibitor and NAD+ binding to poly(ADP-ribose) polymerase as derived from crystal structures and homology modeling. *Biochemistry-Us* **37**, 3893-3900, doi:10.1021/bi972383s (1998).

20 Ha, H. C. & Snyder, S. H. Poly(ADP-ribose) polymerase is a mediator of necrotic cell death by ATP depletion. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13978-13982, doi:DOI 10.1073/pnas.96.24.13978 (1999).

21 Eliasson, M. J. *et al.* Poly(ADP-ribose) polymerase gene disruption renders mice resistant to cerebral ischemia. *Nature medicine* **3**, 1089-1095 (1997).

22 Megnin-Chanet, F., Bollet, M. A. & Hall, J. Targeting poly(ADP-ribose) polymerase activity for cancer therapy. *Cellular and Molecular Life Sciences* **67**, 3649-3662, doi:DOI 10.1007/s00018-010-0490-8 (2010).

23 Martin-Oliva, D. *et al.* Crosstalk between PARP-1 and NF-kappa B modulates the promotion of skin neoplasia. *Oncogene* **23**, 5275-5283, doi:DOI 10.1038/sj.onc.1207696 (2004).

24 Helleday, T., Bryant, H. E. & Schultz, N. Poly(ADP-ribose) polymerase (PARP-1) in homologous recombination and as a target for cancer therapy. *Cell Cycle* **4**, 1176-1178 (2005).

25 Dantzer, F. *et al.* Functional interaction between poly(ADP-ribose) polymerase 2 (PARP-2) and TRF2: PARP activity negatively regulates TRF2. *Mol Cell Biol* **24**, 1595-1607, doi:Doi 10.1128/Mcb.24.4.1595-1607.2004 (2004).

26 Isabelle, M. *et al.* Investigation of PARP-1, PARP-2, and PARG interactomes by affinity-purification mass spectrometry. *Proteome science* **8**, doi:Artn 22
Doi 10.1186/1477-5956-8-22 (2010).

27 Rouleau, M., Patel, A., Hendzel, M. J., Kaufmann, S. H. & Poirier, G. G. PARP inhibition: PARP1 and beyond. *Nat Rev Cancer* **10**, 293-301, doi:Doi 10.1038/Nrc2812 (2010).

28 Yelamos, J. *et al.* PARP-2 deficiency affects the survival of CD4(+)CD8(+) double-positive thymocytes. *Embo Journal* **25**, 4350-4360, doi:DOI 10.1038/sj.emboj.7601301 (2006).

29 Pellicciari, R. *et al.* On the way to selective PARP-2 inhibitors. Design, synthesis, and preliminary evaluation of a series of isoquinolinone derivatives. *Chemmedchem* **3**, 914-923, doi:10.1002/cmdc.200800010 (2008).

30 Kinoshita, T. *et al.* Inhibitor-induced structural change of the active site of human poly(ADP-ribose) polymerase. *Febs Lett* **556**, 43-46, doi:Doi 10.1016/S0014-5793(03)01362-0 (2004).

31 Eltze, T. *et al.* Imidazoquinolinone, imidazopyridine, and isoquinolindione derivatives as novel and potent inhibitors of the poly(ADP-ribose) polymerase (PARP): a comparison with standard PARP inhibitors. *Molecular pharmacology* **74**, 1587-1598, doi:10.1124/mol.108.048751 (2008).

32 Moroni, F. *et al.* Selective PARP-2 inhibitors increase apoptosis in hippocampal slices but protect cortical cells in models of post-ischaemic brain damage. *Brit J Pharmacol* **157**, 854-862, doi:DOI 10.1111/j.1476-5381.2009.00232.x (2009).

33 Ishida, J. *et al.* Discovery of potent and selective PARP-1 and PARP-2 inhibitors: SBDD analysis via a combination of X-ray structural study and homology modeling. *Bioorganic & medicinal chemistry* **14**, 1378-1390, doi:10.1016/j.bmc.2005.09.061 (2006).

34 WITNOTP (Widmer, Armin (Novartis, Basel)).

35 Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph Model* **14**, 33-38, doi:Doi 10.1016/0263-7855(96)00018-5 (1996).

36 Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**, 1545-1614, doi:Doi 10.1002/Jcc.21287 (2009).

37 Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant-Pressure Molecular-Dynamics Algorithms. *J Chem Phys* **101**, 4177-4189, doi:Doi 10.1063/1.467468 (1994).

38 Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J Chem Phys* **103**, 4613-4621, doi:Doi 10.1063/1.470648 (1995).

39 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802, doi:Doi 10.1002/Jcc.20289 (2005).

40 Seeber, M., Cecchini, M., Rao, F., Settanni, G. & Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **23**, 2625-2627, doi:10.1093/bioinformatics/btm378 (2007).

41 Seeber, M. *et al.* Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem* **32**, 1183-1194, doi:10.1002/jcc.21688 (2011).

42 Pascualahuir, J. L., Silla, E. & Tunon, I. Gepol - an Improved Description of Molecular-Surfaces .3. A New Algorithm for the Computation of a Solvent-Excluding Surface. *J Comput Chem* **15**, 1127-1138, doi:DOI 10.1002/jcc.540151009 (1994).

43 Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. Hydrophobicity of Amino-Acid Residues in Globular-Proteins. *Science* **229**, 834-838, doi:DOI 10.1126/science.4023714 (1985).

44 Karlberg, T., Hammarstrom, M., Schutz, P., Svensson, L. & Schuler, H. Crystal structure of the catalytic domain of human PARP2 in complex with PARP inhibitor ABT-888. *Biochemistry-Us* **49**, 1056-1058, doi:10.1021/bi902079y (2010).

45 Steffen, J. D., Brody, J. R., Armen, R. S. & Pascal, J. M. Structural Implications for Selective Targeting of PARPs. *Frontiers in oncology* **3**, 301, doi:10.3389/fonc.2013.00301 (2013).

46 Krivov, S. V. & Karplus, M. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *The journal of physical chemistry. B* **110**, 12689-12698, doi:10.1021/jp060039b (2006).

47 Krivov, S. V. & Karplus, M. Diffusive reaction dynamics on invariant free energy profiles. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13841-13846, doi:DOI 10.1073/pnas.0800228105 (2008).

48 Trott, O. & Olson, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **31**, 455-461, doi:Doi 10.1002/Jcc.21334 (2010).

49 Vanommeslaeghe, K. *et al.* CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem* **31**, 671-690, doi:Doi 10.1002/Jcc.21367 (2010).

50 Yesselman, J. D., Price, D. J., Knight, J. L. & Brooks, C. L., 3rd. MATCH: an atom-typing toolset for molecular mechanics force fields. *J Comput Chem* **33**, 189-202, doi:10.1002/jcc.21963 (2012).

51 Kolb, P. & Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J Med Chem* **49**, 7384-7392, doi:Doi 10.1021/Jm060838i (2006).

52 Huang, D. & Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J Med Chem* **47**, 5791-5797, doi:Doi 10.1021/Jm049726m (2004).

53 Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *Journal of chemical information and modeling* **52**, 1757-1768, doi:10.1021/ci3001277 (2012).

54 Majeux, N., Scarsi, M. & Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins-Structure Function and Genetics* **42**, 256-268, doi:Doi 10.1002/1097-0134(20010201)42:2<256::Aid-Prot130>3.0.Co;2-4 (2001).

55 Majeux, N., Scarsi, M., Apostolakis, J., Ehrhardt, C. & Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins-Structure Function and Genetics* **37**, 88-105, doi:Doi 10.1002/(Sici)1097-0134(19991001)37:1<88::Aid-Prot9>3.0.Co;2-O (1999).

56 Wiedmer, L., Marchand, J. R., Spiliotopoulos, D. & Caflisch, A. *Effective consensus scoring with SEED energy terms* (Department of Biochemistry, University of Zurich, 2014).

57 *RDKit, Open-Source Cheminformatics.*, < http://www.rdkit.org.> (

58 O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *Journal of cheminformatics* **3**, 33, doi:10.1186/1758-2946-3-33 (2011).

59 Xu, M. & Caflisch, A. *In-house adaptation of VINA* (Department of Biochemistry, University of Zurich, 2014).

60 Ruf, A., Mennissier de Murcia, J., de Murcia, G. & Schulz, G. E. Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 7481-7485 (1996).

61 Curtin, N. J. PARP inhibitors for cancer therapy. *Expert reviews in molecular medicine* **7**, 1-20, doi:10.1017/S146239940500904X (2005).

62 Griffin, R. J. *et al.* Resistance-modifying agents. 5. Synthesis and biological properties of quinazolinone inhibitors of the DNA repair enzyme poly(ADP-ribose) polymerase (PARP). *J Med Chem* **41**, 5247-5256, doi:10.1021/jm980273t (1998).

63 White, A. W. *et al.* Resistance-modifying agents. 9. Synthesis and biological properties of benzimidazole inhibitors of the DNA repair enzyme poly(ADP-ribose) polymerase. *J Med Chem* **43**, 4084-4097 (2000).

64 Pellicciari, R. *et al.* Towards new neuroprotective agents: design and synthesis of 4H-thieno[2,3-c] isoquinolin-5-one derivatives as potent PARP-1 inhibitors. *Farmaco* **58**, 851-858 (2003).

65 Perkins, E. *et al.* Novel inhibitors of poly(ADP-ribose) polymerase/PARP1 and PARP2 identified using a cell-based screen in yeast. *Cancer Res* **61**, 4175-4183 (2001).

66 Gubernator, K., Craig, J. A. & Mistry, R. *eMolecules*, <www.emolecules.com> (2005).

67 Nichols, S. E., Baron, R. & McCammon, J. A. On the Use of Molecular Dynamics Receptor Conformations for Virtual Screening. *Computational Drug Discovery and Design* **819**, 93-103, doi:Doi 10.1007/978-1-61779-465-0_7 (2012).

3. REPURPOSING APPROVED DRUGS AND SAFE INHIBITOR

3.1 INTRODUCTION

3.1.1 REPURPOSING MOTIVATION

Sir James Black (1988 Nobel Laureate in Physiology and Medicine) has stated that “[t]he most fruitful basis for the discovery of a new drug is to start with an old drug” [1]. Indeed, not only may a given target biomolecule be implicated in more than one disease, but further drug compounds may bind with high potency to more than one target; of these, targets beyond those originally intended may be relevant to the efficacy of a compound in the context of the disease indication of original development, as well as in further diseases.

Given the time and costs required to reach approval by authorities such as the US Food and Drug Administration (FDA) or European Medicines Agency (EMA), and even to pass through the stages of clinical development, repurposing of compounds found to be safe in humans has the potential to greatly increase the efficacy of the drug development process. Repurposing has already proven successful: as an example, thalidomide is efficacious in the treatment of both leprosy and multiple myeloma [2]. Recently, a complex network approach was used to analyse FDA-approved drugs to hundreds of protein targets, and 23 new drug-target associations were discovered [3], suggesting that this approach to drug finding is likely to meet with success.

3.1.2 PROJECT OBJECTIVES

Capitalising on the scoring function and protocol of the previous Chapter, two drug libraries were docked to ARTD-1: one containing FDA-approved compounds and second corresponding to literature-reported protein kinase inhibitors having successfully passed phase I of clinical trials in humans. The objective was to find compounds from these libraries binding to ARTD-1 with high putative binding affinity.

3.2 COMPOUND LIBRARIES

3.2.1 FDA-APPROVED COMPOUNDS

FDA approved compounds have passed a series of evaluations of their safety. The FDA-approved subset of 3’176 compounds was downloaded from the ZINC database [4]. Compounds of the library were protonated relative to a pH of 7 using obabel [5].

3.2.2 POST-CLINICAL TRIALS PHASE I KINASE INHIBITORS

The ATP substrate of protein kinases shares a number of common chemical moieties with the NAD$^+$ substrate of ARTD-1. For this reason, it was hypothesized that ATP-competitive inhibitors of protein kinases might also bind the ARTD-1 substrate binding pocket. Over 100 small-molecule ATP-competitive inhibitors of protein kinases have passed the first phase of clinical trials and thus have acceptable toxicity profiles over a reasonable dose range. These protein kinase inhibitors were assemble from a thorough review of the literature [6-17], target-based queries of the PubChem Substance and Compound database [18] and DrugBank [19], as well as through searches of the U.S. National Institutes of Health Clinical Trials database [18]. In total, the assembled library contained 119 compounds which were protonated relative to a pH of 7 using obabel [5].

3.3 DOCKING TO ARTD-1

3.3.1 PROTOCOL

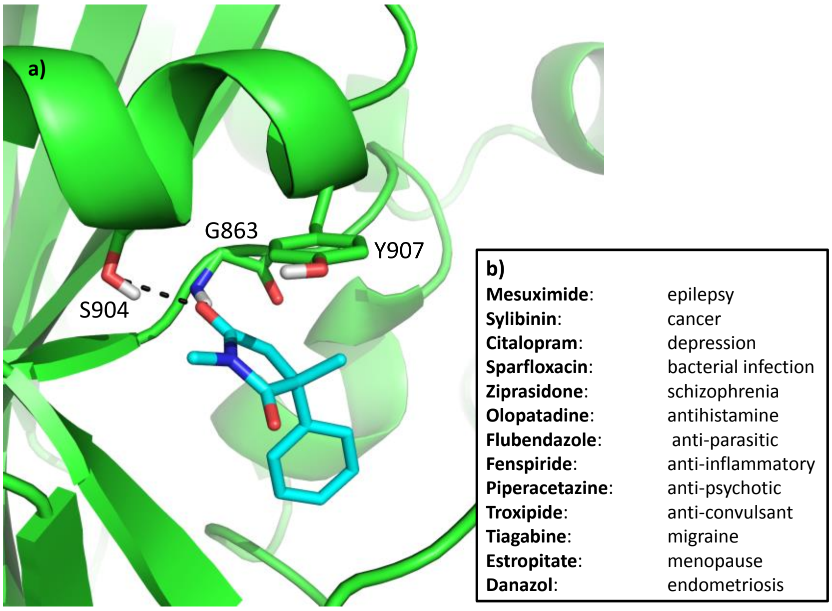
Ligands of each library were docked to the binding site of four protein conformations of ARTD-1 (2 crystal structures and 2 MD-generated conformations as per 2.3.3) using VINA [20] to generate 20 poses per ligand, with filtering for at least one hydrogen bond around the nicotinamide binding site. Docked poses were minimised with CHARMM [21] (by the iterative steepest descent method for 500 steps followed by the conjugate gradient method for 10’000 steps), allowing residues with atoms within 5 Å of the ligand to remain flexible along with the ligand during minimisation. Protein atoms were parametrised with CHARMM27 parameters while ligands were parametrised according to the CHARMM General Force Field [22] with MATCH [23]. The scoring function and cut-off of 2.4.3 and 2.4.4 were applied, with evaluation of terms as per 2.4.2.

3.3.2 RESULTS AND SELECTED COMPOUNDS

Of the original 3’176 compounds forming the FDA-approved compound library, 202 unique compounds met the hydrogen bonding requirements and passed the scoring function cut-off. These were visually inspected and 13 compounds (0.04% of the original library) were selected for experimental validation. These are listed in panel b) of Figure 35, along with their indication of origin, and an example docked pose in panel a).

The post-clinical trials phase I library of kinase inhibitors issued 11 compounds, representing 9% of the original library, selected for experimental validation. These are listed in panel b) of Figure 36, along with their target of origin and an example docked pose in panel a).

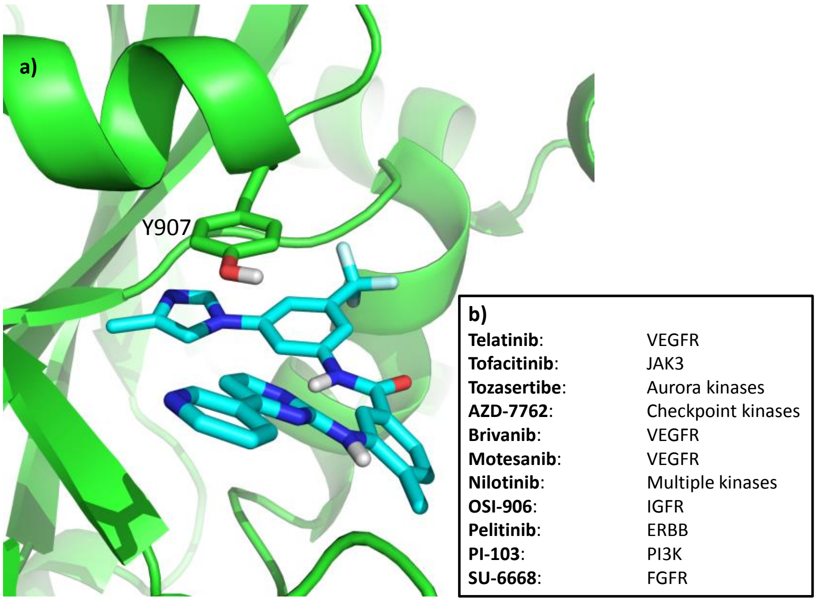
Selected compounds are currently undergoing evaluation at an external service provider.



b)

Mesuximide:	epilepsy
Sylibinin:	cancer
Citalopram:	depression
Sparfloxacin:	bacterial infection
Ziprasidone:	schizophrenia
Olopatadine:	antihistamine
Flubendazole:	anti-parasitic
Fenspiride:	anti-inflammatory
Piperacetazine:	anti-psychotic
Troxipide:	anti-convulsant
Tiagabine:	migraine
Estropitate:	menopause
Danazol:	endometriosis

Figure 35. FDA-approved compounds putatively binding ARTD-1 with high affinity: a) docked pose of mesuximide in the MD-generated protein conformation of Basin II (see 2.3.3), b) list of selected compounds for affinity evaluation, with indication of origin



b)

Telatinib:	VEGFR
Tofacitinib:	JAK3
Tozasertibe:	Aurora kinases
AZD-7762:	Checkpoint kinases
Brivanib:	VEGFR
Motesanib:	VEGFR
Nilotinib:	Multiple kinases
OSI-906:	IGFR
Pelitinib:	ERBB
PI-103:	PI3K
SU-6668:	FGFR

Figure 36. Post-clinical trials phase I kinase inhibitor compounds putatively binding ARTD-1 with high affinity: a) docked pose of nilotinib in the MD-generated protein conformation of Basin I (see 2.3.3), b) list of selected compounds for affinity evaluation, with protein kinase target of origin

3.4 CONCLUSIONS

An effective docking protocol and scoring function has been coupled with an *in silico* screening of known non-toxic compounds. This approach offers an efficient means to repurpose highly-developed compounds quickly and at low cost. Further, it may offer insights into new binding partners of potential relevance to compound efficacy in the context of the disease indication of origin and in the treatment of new indications.

3.5 REFERENCES

1 Chong, C. R. & Sullivan, D. J., Jr. New uses for old drugs. *Nature* **448**, 645-646, doi:10.1038/448645a (2007).

2 Drucker, L. *et al.* Thalidomide down-regulates transcript levels of GC-rich promoter genes in multiple myeloma. *Molecular pharmacology* **64**, 415-420, doi:10.1124/mol.64.2.415 (2003).

3 Keiser, M. J. *et al.* Predicting new molecular targets for known drugs. *Nature* **462**, 175-U148, doi:Doi 10.1038/Nature08506 (2009).

4 Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *Journal of chemical information and modeling* **52**, 1757-1768, doi:10.1021/ci3001277 (2012).

5 O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *Journal of cheminformatics* **3**, 33, doi:10.1186/1758-2946-3-33 (2011).

6 Fabian, M. A. *et al.* A small molecule-kinase interaction map for clinical kinase inhibitors. *Nature biotechnology* **23**, 329-336, doi:10.1038/nbt1068 (2005).

7 Lorusso, P. M. *et al.* Phase I and pharmacodynamic study of the oral MEK inhibitor CI-1040 in patients with advanced malignancies. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **23**, 5281-5293, doi:10.1200/JCO.2005.14.415 (2005).

8 DePinto, W. *et al.* In vitro and in vivo activity of R547: a potent and selective cyclin-dependent kinase inhibitor currently in phase I clinical trials. *Molecular cancer therapeutics* **5**, 2644-2658, doi:10.1158/1535-7163.MCT-06-0355 (2006).

9 Weisberg, E. *et al.* AMN107 (nilotinib): a novel and selective inhibitor of BCR-ABL. *British journal of cancer* **94**, 1765-1769, doi:10.1038/sj.bjc.6603170 (2006).

10 Young, M. A. *et al.* Structure of the kinase domain of an imatinib-resistant Abl mutant in complex with the Aurora kinase inhibitor VX-680. *Cancer Res* **66**, 1007-1014, doi:10.1158/0008-5472.CAN-05-2788 (2006).

11 Carpinelli, P. *et al.* PHA-739358, a potent inhibitor of Aurora kinases with a selective target inhibition profile relevant to cancer. *Molecular cancer therapeutics* **6**, 3158-3168, doi:10.1158/1535-7163.MCT-07-0444 (2007).

12 Yang, J. *et al.* AZD1152, a novel and selective aurora B kinase inhibitor, induces growth arrest, apoptosis, and sensitization for tubulin depolymerizing agent or topoisomerase II inhibitor in human acute leukemia cells in vitro and in vivo. *Blood* **110**, 2034-2040, doi:10.1182/blood-2007-02-073700 (2007).

13 Gonzales, A. J. *et al.* Antitumor activity and pharmacokinetic properties of PF-00299804, a second-generation irreversible pan-erbB receptor tyrosine kinase inhibitor. *Molecular cancer therapeutics* **7**, 1880-1889, doi:10.1158/1535-7163.MCT-07-2232 (2008).

14 Johnson, F. M. *et al.* Phase 1 pharmacokinetic and drug-interaction study of dasatinib in patients with advanced solid tumors. *Cancer* **116**, 1582-1591, doi:10.1002/cncr.24927 (2010).

15 Boss, D. S., Beijnen, J. H. & Schellens, J. H. Clinical experience with aurora kinase inhibitors: a review. *The oncologist* **14**, 780-793, doi:10.1634/theoncologist.2009-0019 (2009).

16 Traxler, P. *et al.* Tyrosine kinase inhibitors: from rational design to clinical trials. *Medicinal research reviews* **21**, 499-512 (2001).

17 Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* **9**, 28-39, doi:10.1038/nrc2559 (2009).

18 <pubchem.ncbi.nlm.nih.gov> (

19 Wishart, D. S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**, D901-906, doi:10.1093/nar/gkm958 (2008).

20 Trott, O. & Olson, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **31**, 455-461, doi:Doi 10.1002/Jcc.21334 (2010).

21 Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**, 1545-1614, doi:Doi 10.1002/Jcc.21287 (2009).

22 Vanommeslaeghe, K. *et al.* CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem* **31**, 671-690, doi:Doi 10.1002/Jcc.21367 (2010).

23 Yesselman, J. D., Price, D. J., Knight, J. L. & Brooks, C. L., 3rd. MATCH: an atom-typing toolset for molecular mechanics force fields. *J Comput Chem* **33**, 189-202, doi:10.1002/jcc.21963 (2012).

4. MACRODOMAIN-CONTAINING PROTEINS ARE NEW MONO-ADP-RIBOSYLHYDROLASES

Rosenthal, F., Feijs, K., Frugier, E., Bonalli, M., Forst, A.H., Imhof, R., Winkler, H.C., Fischer, D., Caflish, A., Hassa, P.O., Lüscher, B. and Hottiger, M. *Nat. Struct. Mol. Biol.*, **2013**, 20(4), 502-509

nature
structural &
molecular biology

Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases

Florian Rosenthal[1,2,6], Karla L H Feijs[3,6], Emilie Frugier[4], Mario Bonalli[1], Alexandra H Forst[3], Ralph Imhof[1], Hans C Winkler[1], David Fischer[5], Amedeo Caflisch[4], Paul O Hassa[1], Bernhard Lüscher[3] & Michael O Hottiger[1]

ADP-ribosylation is an important post-translational protein modification (PTM) that regulates diverse biological processes. ADP-ribosyltransferase diphtheria toxin-like 10 (ARTD10, also known as PARP10) mono-ADP-ribosylates acidic side chains and is one of eighteen ADP-ribosyltransferases that catalyze mono- or poly-ADP-ribosylation of target proteins. Currently, no enzyme is known that reverses ARTD10-catalyzed mono-ADP-ribosylation. Here we report that ARTD10-modified targets are substrates for the macrodomain proteins MacroD1, MacroD2 and C6orf130 from *Homo sapiens* as well as for the macrodomain protein Af1521 from archaebacteria. Structural modeling and mutagenesis of MacroD1 and MacroD2 revealed a common core structure with Asp102 and His106 of MacroD2 implicated in the hydrolytic reaction. Notably, MacroD2 reversed the ARTD10-catalyzed, mono-ADP-ribose–mediated inhibition of glycogen synthase kinase 3β (GSK3β) *in vitro* and in cells, thus underlining the physiological and regulatory importance of mono-ADP-ribosylhydrolase activity. Our results establish macrodomain-containing proteins as mono-ADP-ribosylhydrolases and define a class of enzymes that renders mono-ADP-ribosylation a reversible modification.

ADP-ribosylation is a PTM involved in many biological processes including the regulation of chromatin structure, transcription and DNA repair[1–6]. ARTD10 is a mono-ADP-ribosyltransferase that transfers ADP-ribose to aspartate or glutamate residues by substrate-assisted catalysis[7]. Recently, GSK3β was identified as a new substrate of ARTD10. Mono-ADP-ribosylation noncompetitively inhibits GSK3β kinase activity[8].

Although many ADP-ribosyltransferases (including ARTDs and ADP-ribosyltransferases cholera toxin-like (ARTCs)) that ADP-ribosylate different amino acid acceptor sites have been identified[9], the enzymes able to reverse this modification are largely unknown. Irreversible ADP-ribosylation is highly detrimental and causes embryonic lethality[10,11]. In humans, the deficiency of an ADP-ribose hydrolase is the cause of fatal glutamyl ribose 5-phosphate storage disease[12]. Therefore, enzymes that remove ADP-ribose modifications must exist. The mammalian hydrolases characterized so far include the mono-ADP-ribosylarginine hydrolase 1 (ARH1) as well as poly-ADP-ribose glycohydrolases (PARG and ARH3). Whereas ARH1 is the only hydrolase that specifically removes mono-ADP-ribose from arginine residues, ARH3 and PARG hydrolyze the O-glycosidic ribose-ribose 1-2′ bonds within ADP-ribose polymers[13,14]. In contrast, enzymes catalyzing the removal of specific mono-ADP-ribose marks of modified aspartates or glutamates, such as those synthesized by ARTD10, are currently unknown (**Supplementary Fig. 1a**).

Macrodomains are a family of evolutionarily conserved proteins that bind mono- or poly-ADP-ribose (PAR), poly(A) or *O*-acetyl-ADP-ribose (OAADPr)[15–18]. Macrodomain proteins are involved in diverse cellular processes[15] and have been implicated in transcriptional regulation[19–22], chromatin remodeling[23,24] and developmental processes as well as in B-cell lymphomagenesis[25–27]. Macrodomain-containing proteins localize to the nucleus (for example, MacroD1v2) or to mitochondria (MacroD1v1) or are found in the cytoplasm (MacroD1v2, MacroD2 or C6orf130)[18]. In addition to the binding of ADP-ribose, the human MacroD1, MacroD2 and C6orf130 possess C2- or C3-specific OAADPr deacetylase activity[28,29] and weak C1-specific phosphatase activity toward ADP-ribose-1″-phosphate (Appr-1″-p), a product of ARTD18 (TPT1) and cyclic nucleotide phosphodiesterases[17,30,31]. On the basis of structural and functional analyses, several residues in the active centers of macrodomains were identified that participate in the catalytic mechanism[28,29]. Notably, the catalytically important residues of MacroD1 are not conserved in C6orf130, which indicates that sequence variation within the macrodomain family allows a different set of catalytic residues to perform OAADPr hydrolysis[29].

Here we test the hypothesis that macrodomain-containing proteins possess ADP-ribosylhydrolase activity and thereby to characterize the missing mono-ADP-ribosylhydrolases. We provide evidence that the human proteins MacroD1, MacroD2 and C6orf130 as well as the archaebacterial macrodomain Af1521 are able to hydrolyze ARTD10-catalyzed mono-ADP-ribosylation. Notably, MacroD2 rendered the inhibitory effect of GSK3β ADP-ribosylation reversible. Treatment with MacroD2 removed the ADP-ribose moiety from GSK3β, which was sufficient to restore kinase activity *in vitro* and in cells. These

---

[1]Institute of Veterinary Biochemistry and Molecular Biology, University of Zurich, Zurich, Switzerland. [2]Life Science Zurich Graduate School, University of Zurich, Zurich, Switzerland. [3]Institute of Biochemistry and Molecular Biology, Rheinisch-Westfälische Technische Hochschule Aachen, Aachen, Germany. [4]Department of Biochemistry, University of Zurich, Zurich, Switzerland. [5]Functional Genomics Center Zurich, University of Zurich, Zurich, Switzerland. [6]These authors contributed equally to this work. Correspondence should be addressed to B.L. (luescher@rwth-aachen.de) or M.O.H. (hottiger@vetbio.uzh.ch).

Received 12 September 2012; accepted 17 January 2013; published online 10 March 2013; doi:10.1038/nsmb.2521

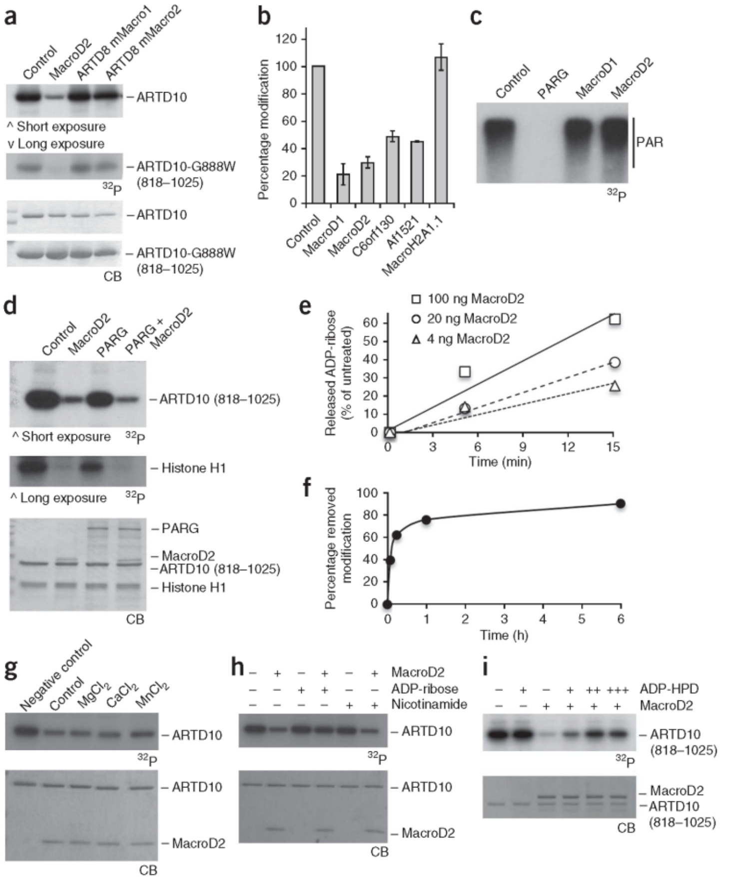


Figure 1 MacroD1, MacroD2, C6orf130 and Af1521 hydrolyze mono-ADP-ribose modifications. (**a**,**b**) Protein mono-ADP-ribosylhydrolase activity of MacroD1, MacroD2, C6orf130, mouse ARTD8 (mMacro1 and mMacro2), macroH2A1.1 and Af1521. Auto-ADP-ribosylated ARTD10 de-ADP-ribosylated by the indicated proteins is shown on SDS-PAGE with Coomassie blue (CB) staining or autoradiography (^{32}P). Shown are representative blots and quantification from two independent experiments, averaged and normalized to the untreated control ($n = 2$; mean ± range). (**c**) Hydrolase activity assays showing that MacroD-like macrodomains have no activity toward polymers of ADP-ribose, whereas PARG has strong activity. Auto-poly-ADP-ribosylated ARTD1 (with 160 μM radioactive NAD$^+$ to induce poly-ADP-ribosylation) was used with the indicated hydrolases. (**d**) Activity assays as in **a**, showing that PARG cannot hydrolyze ADP-ribose from mono-ADP-ribosylated ARTD10 or from histones. (**e**) Concentration-dependent removal of the mono-ADP-ribose from ARTD10 (818–1025) by MacroD2 under nonsaturating conditions. (**f**) Time course of MacroD2 activity toward mono-ADP-ribosylated ARTD10 catalytic domains. (**g**) Identification of cofactors (1 mM each) for ADP-ribose hydrolysis by MacroD2. (**h**) Effect of ADP-ribose and nicotinamide (40 μM each) on MacroD2-mediated de-ADP-ribosylation of ARTD10. (**i**) The ADP-ribose analog ADP-HPD inhibits MacroD2 in a concentration-dependent manner (4, 40 and 400 μM).

data highlight the important physiological function of endogenous mono-ADP-ribosylation for intracellular signaling and regulatory processes.

RESULTS

Macrodomain proteins are mono-ADP-ribosylhydrolases

To investigate whether macrodomain-containing proteins are able to release the mono-ADP-ribose moiety from ARTD10-modified target proteins, we incubated different macrodomains with *in vitro*–radiolabeled mono-ADP-ribosylated ARTD10 (full length or the catalytic domain, consisting of residues 818–1025) as a substrate (**Fig. 1a**,**b** and **Supplementary Fig. 1b**,**c**). Human MacroD1, MacroD2 and C6orf130 robustly hydrolyzed the mono-ADP-ribosyl linkage of modified full-length ARTD10 or the ARTD10 catalytic domain. Notably, the structurally related archaebacterial macrodomain protein Af1521 was also active, whereas neither the human histone variant MacroH2A1.1 nor the mouse macrodomains 1 or 2 of ARTD8 (PARP14) were able to remove mono-ADP-ribose under the tested conditions (**Fig. 1a**,**b** and **Supplementary Fig. 1b**,**c**). Pulldown experiments revealed that mono-ADP-ribosylhydrolase activity of the tested macrodomains correlated with the ability to bind mono-ADP-ribosylated ARTD10 (**Supplementary Fig. 1d**). In contrast to their activity toward mono-ADP-ribosylated substrates, MacroD1 and MacroD2 were each completely inactive toward PAR synthesized by ARTD1 (PARP1) in the presence of high NAD$^+$ concentrations, whereas the known PAR hydrolase PARG exhibited strong activity toward this substrate (**Fig. 1c**). However, PARG was not able to completely remove all ADP-ribose modifications from automodified ARTD1, even when tested under low NAD$^+$ concentrations (resulting in mono- and oligo-ADP-ribosylation; **Supplementary Fig. 1e** and as suggested in ref. 13). The nature of the modified protein did not affect PARG-dependent hydrolysis, because histones modified by ARTD1 under low NAD$^+$ concentrations were also not completely demodified (**Supplementary Fig. 1f**). Notably, prior PARG treatment rendered ARTD1, histone H1 and core histones at least partially susceptible to hydrolysis by MacroD2 (**Supplementary Fig. 1f**,**g**), which suggests that the acceptor residue and/or the linkage (C1 versus C2 or C3) between the ADP-ribose moiety and the acceptor residue are critical for the newly identified enzymatic activity. Consequently, PARG treatment probably generates mono-ADP-ribosylated residues that serve as substrates for hydrolysis by MacroD2. Comparably, PARG was inactive toward ARTD10-catalyzed mono-ADP-ribosylated histone H1, whereas MacroD2 demodified this substrate, irrespective of PARG treatment (**Fig. 1d**). These results suggested that MacroD2 is able to release mono-ADP-ribose from acceptor proteins but is inactive toward polymers or oligomers of ADP-ribose.

To biochemically characterize the enzymatic reaction catalyzed by MacroD2, we performed concentration- and time-dependent experiments. MacroD2 efficiently removed ADP-ribose modifications from

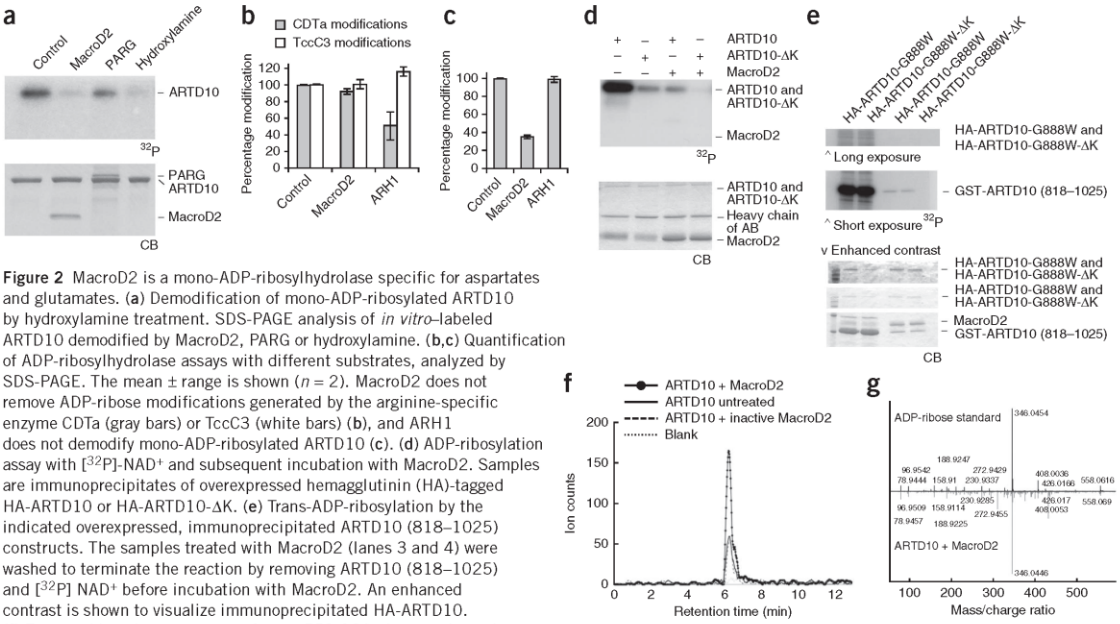


Figure 2 MacroD2 is a mono-ADP-ribosylhydrolase specific for aspartates and glutamates. (**a**) Demodification of mono-ADP-ribosylated ARTD10 by hydroxylamine treatment. SDS-PAGE analysis of *in vitro*–labeled ARTD10 demodified by MacroD2, PARG or hydroxylamine. (**b**,**c**) Quantification of ADP-ribosylhydrolase assays with different substrates, analyzed by SDS-PAGE. The mean ± range is shown ($n = 2$). MacroD2 does not remove ADP-ribose modifications generated by the arginine-specific enzyme CDTa (gray bars) or TccC3 (white bars) (**b**), and ARH1 does not demodify mono-ADP-ribosylated ARTD10 (**c**). (**d**) ADP-ribosylation assay with [^{32}P]-NAD$^+$ and subsequent incubation with MacroD2. Samples are immunoprecipitates of overexpressed hemagglutinin (HA)-tagged HA-ARTD10 or HA-ARTD10-ΔK. (**e**) Trans-ADP-ribosylation by the indicated overexpressed, immunoprecipitated ARTD10 (818–1025) constructs. The samples treated with MacroD2 (lanes 3 and 4) were washed to terminate the reaction by removing ARTD10 (818–1025) and [^{32}P] NAD$^+$ before incubation with MacroD2. An enhanced contrast is shown to visualize immunoprecipitated HA-ARTD10. (**f**) Extracted ion chromatograms of mass 558.064, the monoisotopic mass of the ADP-ribose [M-H] adduct. The elution time window of ADP-ribose was between 6 and 7 min. (**g**) MS-MS spectrum of mass 558.064. Typical mass fragments of ADP-ribose as seen in the ADP-ribose standard (top spectrum) are present in the MS-MS spectrum of ARTD10 + MacroD2 samples (bottom spectrum).

ARTD10 (818–1025) in a concentration-dependent manner (**Fig. 1e**). Furthermore, mono-ADP-ribose hydrolysis by MacroD2 was time dependent and removed >60% of the modifications within 15 min (**Fig. 1f**). Further characterization of the enzymatic activity revealed that MacroD2 activity was not markedly affected by the addition of magnesium, calcium or manganese as cofactor (**Fig. 1g**). Notably, MacroD2 activity was inhibited by the addition of the ADP-ribose analog adenosine 5′-diphosphate (hydroxymethyl) pyrrolidinediol (ADP-HPD) and by ADP-ribose itself in a concentration-dependent manner, whereas addition of nicotinamide did not influence the mono-ADP-ribosylhydrolase activity (**Fig. 1h**,**i**). These observations indicated that free ADP-ribose is able to inhibit the mono-ADP-ribose hydrolyzing activity by competing with protein-linked mono-ADP-ribose for binding to the active site. Together, these experiments thus defined MacroD1, MacroD2, C6orf130 and Af1521 as new specific mono-ADP-ribosylhydrolases.

MacroD2 is a mono-ADP-ribosylhydrolase for acidic residues

To define the specificity of MacroD2 mono-ADP-ribosylhydrolase activity, we characterized and analyzed different substrates. Hydroxylamine treatment was reported to remove ADP-ribose from glutamate and arginine residues[32]. Treatment of ARTD10 with hydroxylamine for 60 min at 37 °C released the modification of the enzyme, as already observed earlier (**Fig. 2a** and ref. 7), which suggests that MacroD2 is able to hydrolyze mono-ADP-ribose from acidic acceptor sites or arginine. To further investigate whether MacroD2 removed ADP-ribose from arginine acceptor amino acids, the arginine-specific ADP-ribosyltransferase CDTa was used[33]. MacroD2 showed no activity toward actin modified by CDTa (**Fig. 2b**), which indicates that modified arginine residues cannot be hydrolyzed by MacroD2. In contrast, the arginine-specific hydrolase ARH1 was able to remove the ADP-ribose from arginine-modified β- or γ-actin but was not able to remove ADP-ribose from ARTD10 (**Fig. 2c**), which confirms that ARTD10 is not modified at arginine residues. Actin modified at threonine residues by the threonine-specific transferase TccC3 (ref. 34) could not be demodified by either ARH1 or MacroD2 (**Fig. 2b**). To exclude that lysine residues were modified by ARTD10 and consecutively demodified by MacroD2, we mutated all ARTD10 lysine residues to arginines. This mutant (ARTD10-ΔK) showed reduced ADP-ribosylation activity toward both itself and GSK3β, a newly identified ARTD10 target (**Fig. 2d** and **Supplementary Fig. 2a**), which indicated that mutation of these lysine residues interferes with the enzymatic activity of ARTD10. However, the inactive mutant ARTD10-G888W-ΔK could be modified in *trans* by the catalytic domain of ARTD10 to a comparable extent as ARTD10-G888W (**Fig. 2e**), thus implying that lysines are not the acceptor sites. Of note, ARTD10-ΔK and ARTD10-G888W-ΔK were still demodified by MacroD2 to a comparable extent as were wild-type ARTD10 or ARTD10-G888W, respectively (**Fig. 2d**,**e**), which suggests that the same residues (for example, aspartates or glutamates) are automodified in ARTD10-ΔK and wild-type ARTD10 and consecutively demodified by MacroD2. In conclusion, these results suggested that MacroD2 probably releases ADP-ribose from ADP-ribosylated acidic residues.

To confirm that MacroD2 removes ADP-ribose from its target protein, the reaction products were analyzed by LC-MS and HPLC. Upon incubation of ARTD10 with MacroD2, a product that eluted at the same time as the ADP-ribose standard and had the expected mass of 558.064 Da was detected, thus showing that MacroD2 indeed removed ADP-ribose from ARTD10 (**Fig. 2f**,**g** and **Supplementary Fig. 2b**–**d**).

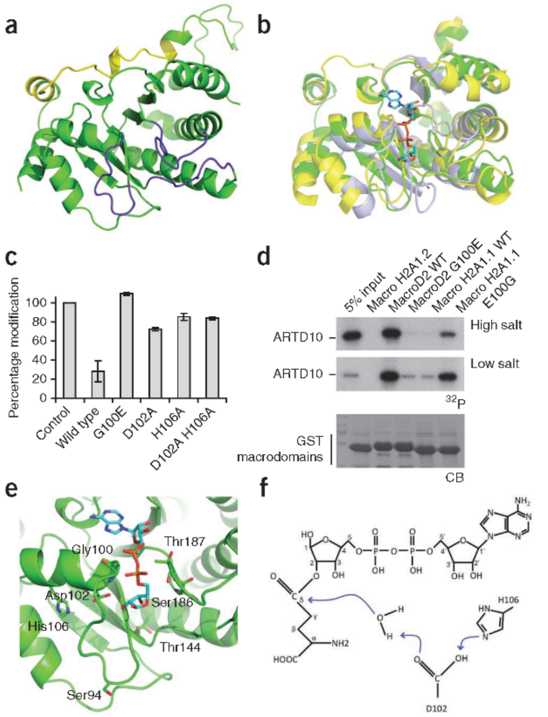


Figure 3 Structural and mutational analysis of the mono-ADP-ribosylhydrolase activity of MacroD2. (**a**) Structure of MacroD2 modeled on the basis of the PDB 2X47 crystal structure of MacroD1. Residues highlighted in yellow represent those having a low r.m.s. fluctuation value as compared to the crystal-structure *B* factors. The primary macrodomain binding-site loops are marked in violet. (**b**) Overlay of the MacroD1 (yellow), MacroD2 (green) and C6orf130 (violet) structures with the ADP-ribose product in the binding pocket. (**c**) Mutational analysis of MacroD2 highlights residues implicated in enzymatic activity toward mono-ADP-ribosylated ARTD10 (818–1025). Quantifications from blots of two independent experiments are shown normalized to the untreated control (*n* = 2; mean ± range). (**d**) ADP-ribose binding to wild-type (WT) and mutant macrodomain proteins, assessed with histidine- and glutathione-S-transferase (GST)-tagged, recombinantly expressed macrodomain proteins immobilized on glutathione or nickel Sepharose and incubated with automodified ARTD10. (**e**) Structural model of MacroD2 with the mutated residues highlighted. (**f**) Model for MacroD2 protein–catalyzed hydrolysis of mono-ADP-ribosylated glutamate residues.

Modeling of MacroD2 and of residues implicated in catalysis

Homology modeling and atomistic simulations were carried out to shed light on the hydrolysis reaction mechanism. First, the three-dimensional structure of MacroD2 was modeled by using the X-ray structure of MacroD1 as a template (**Fig. 3a**). Except for their flexible loops, the human MacroD1, MacroD2 and C6orf130 structures are highly similar, as emerges from their superposition (**Fig. 3b**). This structural likeness, along with their similar enzymatic activities, suggested a conserved mode of action.

Second, the ADP-ribose product of the hydrolysis reaction was automatically docked to MacroD2 and followed by multiple explicit solvent molecular dynamics simulations of the complex to validate the binding mode. The mono-ADP-ribosylhydrolase activity was then studied by mutational analysis of MacroD2 and by comparison with existing mutants of MacroD1 (ref. 29) because these two macrodomain-containing proteins are closely related.

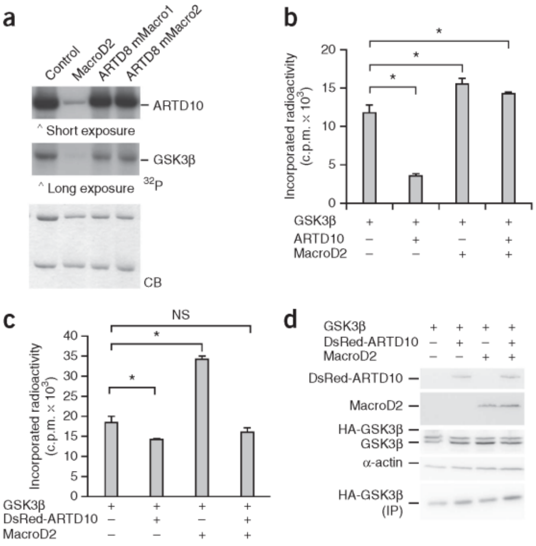
To confirm that the hydrolase activity of MacroD2 is dependent specifically on the macrodomain, we mutated the conserved glycine at position 100 of MacroD2 (**Supplementary Fig. 3a**) to a glutamate, which is predicted to block the ADP-ribose–binding site in macrodomain proteins[35], or to an isoleucine. The resulting MacroD2 mutants lacked hydrolase and ADP-ribose–binding activity under the same assay conditions, which provided evidence that the macrodomain is responsible for the catalytic activity toward ARTD10-mediated mono-ADP-ribosylation and that the interaction with ADP-ribose is specific (**Fig. 3c**,**d** and data not shown). Notably, MacroH2A1.1 contains a glutamate at the corresponding position 225 instead of a glycine (**Supplementary Fig. 3a**), which possibly explains its inactivity due to its inability to bind mono-ADP-ribosylated ARTD10. However, although mutation of Glu225 of MacroH2A1.1 to a glycine resulted in a gain of binding, it did not restore its enzymatic activity (**Supplementary Fig. 3b**), thus indicating that additional residues are important for the activity.

Multiple explicit solvent molecular dynamics simulations of the MacroD2–ADP-ribose complex, together with defining conserved residues between MacroD1, MacroD2 and C6orf130, was used to propose site-specific mutants of MacroD2. Of particular note were Asp102 and its buried neighbor His106, which by modeling were predicted to be located near the 2- and 3-hydroxyl groups of the distal ribose (**Fig. 3e**). Mutational analysis, guided by the *in silico* predictions, indicated partial involvement in catalysis of Asp102 and His106 of MacroD2 as well as Asp184 and His188 of MacroD1 (**Fig. 3c** and **Supplementary Fig. 3c**). Pulldown experiments revealed that the MacroD2 mutants were still able to bind mono-ADP-ribosylated ARTD10 under the conditions tested for their enzymatic activity but that increasing the salt concentration to 500 mM reduced their affinity to the substrate to some extent (**Supplementary Fig. 3d**). On the basis of our findings, a model emerges for MacroD2-catalyzed hydrolysis of mono-ADP-ribosylated aspartate or glutamate residues (**Fig. 3f**). In this model, which is similar to the one previously suggested for the hydrolysis of OAADPr by MacroD1 (ref. 28), Asp102 or Asp184 acts as a general base that deprotonates a water molecule, which then acts as a nucleophile to attack the carbonyl carbon. We note that our mutagenesis data do not exclude participation of additional residues and/or transition-state stabilization due to induced fit.

MacroD2 regulates GSK3β function *in vitro* and in cells

GSK3β is a key regulator in processes ranging from cell structure and survival to diseases such as Alzheimer's disease, cancer and diabetes[36], and it was recently identified in a screen for ARTD10 target proteins (ref. 8 and **Supplementary Fig. 2a**). To address whether the ARTD10-catalyzed ADP-ribosylation of GSK3β is reversible (in addition to the observed ribosylation of ARTD10 and histones (**Figs. 1a**,**b**,**d** and **2e**)), the ADP-ribosylhydrolase activity of MacroD2 on GSK3β was analyzed. Notably, MacroD2 removed the ADP-ribosylation from both ARTD10 and GSK3β *in vitro* (**Fig. 4a**). These findings support the notion that ARTD10-mediated mono-ADP-ribosylation of target proteins in general is a reversible PTM. To test whether removal of the inhibitory mono-ADP-ribose by MacroD2 is sufficient to restore GSK3β kinase activity, mono-ADP-ribosylated GSK3β was demodified by MacroD2 and consecutively used in kinase assays with a primed peptide substrate. Although the ADP-ribosylated protein showed little *in vitro* kinase activity compared to the control, de-ADP-ribosylation restored the GSK3β activity (**Fig. 4b**).

Figure 4 MacroD2 functionally and specifically regulates mono-ADP-ribosylation of GSK3β *in vitro* and in cells. MacroD2 but not the catalytically inactive macrodomains of ARTD8 removes mono-ADP-ribose from ARTD10 and GSK3β. (**a**) Coomassie blue (CB) staining and autoradiography (^{32}P) results of *in vitro* assays with tandem-affinity purification (TAP)-tagged ARTD10 and GST-GSK3β coupled to beads. (**b**) Scintillation counting of kinase assays using [γ-^{32}P]ATP and substrate peptide with modified or demodified GST-GSK3β. (**c**) Scintillation counting results as in **b** with HA-GSK3β from U2OS cells expressing DsRed-ARTD10, GFP-MacroD2 or both. (**d**) Input blot showing expression of DsRed-ARTD10, GFP-MacroD2 and HA-GSK3β in U2OS cells and immunoprecipitation efficiency of HA-GSK3β (lower blot). Data are represented as mean ± s.d. of at least triplicate measurements from representative experiments. $^{*}P < 0.05$; NS, not significant by two-tailed Student's *t* test.



To address whether MacroD2 also reversed ARTD10-mediated ADP-ribosylation in cells, HA-GSK3β was coexpressed with the fluorescently tagged DsRed-ARTD10 alone or together with GFP-MacroD2, and subsequently immunoprecipitated and included in a kinase assay (**Fig. 4c**). The presence of MacroD2 enhanced the kinase activity of GSK3β, which suggested that MacroD2 antagonized the ARTD10-mediated inactivation of GSK3β without affecting protein expression (**Fig. 4d**). These data indicated that MacroD2 hydrolyzes mono-ADP-ribosyl linkages in cells, thus rendering mono-ADP-ribosylation a dynamic modification that can regulate the activation or inactivation of proteins such as GSK3β.

DISCUSSION

The biochemical experiments, mutational analyses and structural predictions presented here suggest that certain macrodomain-containing proteins such as MacroD1, MacroD2, C6orf130 or Af1521, but neither the macrodomains 1 and 2 of ARTD8 nor macroH2A1.1, are bona fide protein mono-ADP-ribosylhydrolases. These results thus fill a major gap in understanding of the ADP-ribosylation cycle and define these proteins as new mono-ADP-ribosylhydrolases that reverse the PTM catalyzed by ARTD10.

The Appr-1″-p phosphatase activity and the OAADPr-hydrolyzing activity of macrodomain proteins have been previously described and studied[28–30]. OAADPr represents an ADP-ribose that is O-acetylated at the C2 or C3 atom, whereas the PAR chain features C1 linkages[37–39]. The OAADPr hydrolysis by MacroD-like macrodomains suggests that these proteins hydrolyze ADP-riboses with C2 or C3 linkages at the proximal ribose. In contrast, the lack of MacroD1 or MacroD2 and C6orf130 activity toward PAR suggests that the glycosidic (ribose-ribose 1′-2′) linkages at the C1 atom of ADP-ribose polymers are not attacked, whereas these bonds within PAR are efficiently degraded by PARG and ARH3. The inefficient hydrolysis of ADP-ribose from ARTD10 by PARG suggests that the C2 or C3 atoms form the glycosidic bond to a glutamate or aspartate (similar to the linkage in OAADPr) rather than to the C1 atom as in PAR. Alternatively, MacroDs might recognize not only the ADP-ribose but also parts of the modified target protein. Moreover, the inability of PARG and MacroD1 or MacroD2 to completely remove the protein-linked ADP-ribose unit from ARTD1 suggests that at least some of the modifications catalyzed by ARTD1 are linked to yet another acceptor site.

The MacroD2-mediated reaction seems to be very efficient because at a 1:10 ratio (MacroD2/ARTD10) more than 60% of the modification synthesized by ARTD10 was removed within 15 min *in vitro*. Owing to the low rates of enzymatic mono-ADP-ribosylation, it was not possible to generate sufficient amounts of mono-ADP-ribosylated MacroD2 substrate that would allow substrate-saturated conditions to exactly determine V_{max} and K_m. In addition, it has been previously shown that ARTD10 is ADP-ribosylated at multiple sites[7]. Because the MacroD2 affinity for these different acceptor sites probably varies, it is impossible to determine exact kinetics with such a substrate. However, chemical synthesis of defined mono-ADP-ribosylated peptides, which could serve as substrates, is currently not possible.

The activity of MacroD-like proteins toward other acceptor sites remains to be tested, but the different chemical nature of these linkages probably requires specific enzymes for the different acceptor sites (for example, ARH1). Furthermore, the specificity may additionally be determined by protein-protein interactions. However, as MacroD2 is expressed in the cytoplasm and MacroD1 exhibits nuclear and mitochondrial localization[18], these two proteins are likely to encounter different protein substrates but catalyze the same mono-ADP-ribosylhydrolysis reaction. ARTD10 and most of the other mono-ARTDs are mainly localized in the cytoplasm[40]. ARTD10-modified target proteins thus require cytoplasmic MacroD proteins for demodification.

On the basis of crystal structures, several residues in MacroD1 and C6orf130 were mutated in other studies to identify side chains involved in OAADPr hydrolysis[28,29]. Notably, we find that corresponding residues are implicated in MacroD1- and MacroD2-mediated hydrolysis of residues ADP-ribosylated by ARTD10. Mutating any of the putative catalytic residues (Asp102 and His106 of MacroD2) individually or together resulted in only a partial loss of activity, which indicates that other residues might compensate for the single mutations, to a certain extent. This observation is further supported by the fact that C6orf130 is active even though the residues identified in MacroD1 and MacroD2 are not conserved in C6orf130 (**Supplementary Fig. 2a**), which again points at sequence variation allowing different sets of residues to confer hydrolase activity. The residues Ser35 and Asp125, crucial for the hydrolysis of OAADPr by C6orf130, would indeed be strong candidates[28].

The removal of ADP-ribose from GSK3β is sufficient to restore kinase activity, which indicates that mono-ADP-ribosylation is a dynamic PTM that directly influences the catalytic activity of its substrates in a reversible manner. The hydrolases identified here represent the missing link in the regulatory network formed by mono-ADP-ribosylation, which may prove highly important for diverse signaling networks as implied by the diversity of ARTD10 substrates identified[8].

In summary, the findings presented here define the macrodomain-containing proteins MacroD1, MacroD2 and C6orf130 as protein mono-ADP-ribosylhydrolases and thus establish mono-ADP-ribosylation of acidic residues by ARTD10 as a reversible PTM. The MacroD-like proteins unite specific ADP-ribose binding with ADP-ribose degradation and thereby define new players in ADP-ribose metabolism and function. MacroD-like hydrolases form the functional antagonists of intracellular mono-ADP-ribosyltransferases. Notably, substrates of the mono-ADP-ribosyltransferase ARTD10 include many kinases, which may thus be activated and inactivated by the opposite activities of MacroD2 and ARTD10.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS
We are grateful to I. Ahel (Paterson Institute for Cancer Research, Manchester, UK) and M. Neuvonen (Institute of Biotechnology, University of Helsinki, Helsinki, Finland) for MacroD1 and MacroD2 constructs and J. Moss (National Institutes of Health, Bethesda, Maryland, USA) for the ARH1 construct. F. Althaus (University of Zurich, Zurich, Switzerland) and K. Aktories (University of Freiburg, Freiburg, Germany) are acknowledged for generously providing a baculovirus expressing PARG and the CDTa and TccC3 enzymes together with purified actin, respectively. F. Freimoser (University of Zurich, Zurich, Switzerland) provided editorial assistance and critical input during the writing and M. Fey technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft DFG (LU 466/15-1) to B.L. and in part by the University of Zurich (Forschungskredit 54041205 to E.F.), the Swiss National Science Foundation (SNF-31003A_125190 (to P.O.H.) and 31-122421 (to M.O.H.)) and the Kanton of Zurich (to M.O.H.).

AUTHOR CONTRIBUTIONS

F.R. and P.O.H. performed experiments with MacroD1, MacroD2 and other macrodomains, and MS analysis was done together with D.F.; K.L.H.F. analyzed the influence of MacroD2 on GSK3β; E.F. performed the computer modeling; E.F. and A.C. analyzed the modeling and simulation results; M.B. identified MacroD2 as an ADP-ribosylhydrolase; A.H.F. cloned and purified mMacro1 and mMacro2; R.I. performed the HPLC analysis; H.C.W. compared the macro sequences; A.C., P.O.H., B.L. and M.O.H. supervised the work. B.L. and M.O.H. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Erener, S. *et al.* ARTD1 deletion causes increased hepatic lipid accumulation in mice fed a high-fat diet and impairs adipocyte function and differentiation. *FASEB J.* **26**, 2631–2638 (2012).
2. Altmeyer, M. & Hottiger, M.O. Poly(ADP-ribose) polymerase 1 at the crossroad of metabolic stress and inflammation in aging. *Aging* **1**, 458–469 (2009).
3. Hassa, P.O., Haenni, S., Elser, M. & Hottiger, M.O. Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? *Microbiol. Mol. Biol. Rev.* **70**, 789–829 (2006).
4. Telli, M.L. PARP inhibitors in cancer: moving beyond BRCA. *Lancet Oncol.* **12**, 827–828 (2011).
5. Curtin, N.J. PARP inhibitors for cancer therapy. *Expert Rev. Mol. Med.* **7**, 1–20 (2005).
6. Schreiber, V., Dantzer, F., Ame, J.-C. & De Murcia, G. Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell Biol.* **7**, 517–528 (2006).
7. Kleine, H. *et al.* Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation. *Mol. Cell* **32**, 57–69 (2008).
8. Feijs, K.L.H. *et al.* ARTD10 substrate identification on protein microarrays: regulation of GSK3β by mono-ADP-ribosylation. *Cell Commun. Signal* published online, doi:10.1186/1478-811X-11-5 (19 January 2013).
9. Hottiger, M.O., Hassa, P.O., Lüscher, B., Schüler, H. & Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* **35**, 208–219 (2010).
10. Hanai, S. *et al.* Loss of poly(ADP-ribose) glycohydrolase causes progressive neurodegeneration in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **101**, 82–86 (2004).
11. Koh, D.W. *et al.* Failure to degrade poly(ADP-ribose) causes increased sensitivity to cytotoxicity and early embryonic lethality. *Proc. Natl. Acad. Sci. USA* **101**, 17699–17704 (2004).
12. Williams, J.C., Chambers, J.P. & Liehr, J.G. Glutamyl ribose 5-phosphate storage disease. A hereditary defect in the degradation of poly(ADP-ribosylated) proteins. *J. Biol. Chem.* **259**, 1037–1042 (1984).
13. Slade, D. *et al.* The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. *Nature* **477**, 616–620 (2011).
14. Dunstan, M.S. *et al.* Structure and mechanism of a canonical poly(ADP-ribose) glycohydrolase. *Nat. Commun.* **3**, 878 (2012).
15. Han, W., Li, X. & Fu, X. The macro domain protein family: structure, functions, and their potential therapeutic implications. *Mutat. Res.* **727**, 86–103 (2011).
16. Kim, I.K. *et al.* Structure of mammalian poly(ADP-ribose) glycohydrolase reveals a flexible tyrosine clasp as a substrate-binding element. *Nat. Struct. Mol. Biol.* **19**, 653–656 (2012).
17. Karras, G.I. *et al.* The macro domain is an ADP-ribose binding module. *EMBO J.* **24**, 1911–1920 (2005).
18. Neuvonen, M. & Ahola, T. Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. *J. Mol. Biol.* **385**, 212–225 (2009).
19. Angelov, D. *et al.* The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. *Mol. Cell* **11**, 1033–1041 (2003).
20. Buschbeck, M. *et al.* The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat. Struct. Mol. Biol.* **16**, 1074–1079 (2009).
21. Changolkar, L.N. *et al.* Developmental changes in histone macroH2A1-mediated gene regulation. *Mol. Cell Biol.* **27**, 2758–2764 (2007).
22. Goenka, S., Cho, S.H. & Boothby, M. Collaborator of Stat6 (CoaSt6)-associated poly(ADP-ribose) polymerase activity modulates Stat6-dependent gene transcription. *J. Biol. Chem.* **282**, 18732–18739 (2007).
23. Ahel, D. *et al.* Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* **325**, 1240–1243 (2009).
24. Gottschalk, A.J. *et al.* Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proc. Natl. Acad. Sci. USA* **106**, 13770–13774 (2009).
25. Aguiar, R.C., Takeyama, K., He, C., Kreinbrink, K. & Shipp, M.A. B-aggressive lymphoma family proteins have unique domains that modulate transcription and exhibit poly(ADP-ribose) polymerase activity. *J. Biol. Chem.* **280**, 33756–33765 (2005).
26. Aguiar, R.C. *et al.* BAL is a novel risk-related gene in diffuse large B-cell lymphomas that enhances cellular migration. *Blood* **96**, 4328–4334 (2000).
27. Cho, S.H. *et al.* Glycolytic rate and lymphomagenesis depend on PARP14, an ADP ribosyltransferase of the B aggressive lymphoma (BAL) family. *Proc. Natl. Acad. Sci. USA* **108**, 15972–15977 (2011).
28. Chen, D. *et al.* Identification of macrodomain proteins as novel O-acetyl-ADP-ribose deacetylases. *J. Biol. Chem.* **286**, 13261–13271 (2011).
29. Peterson, F.C. *et al.* Orphan macrodomain protein (human C6orf130) is an O-acyl-ADP-ribose deacylase: solution structure and catalytic properties. *J. Biol. Chem.* **286**, 35955–35965 (2011).
30. Hofmann, A. *et al.* Structure and mechanism of activity of the cyclic phosphodiesterase of Appr>p, a product of the tRNA splicing reaction. *EMBO J.* **19**, 6207–6217 (2000).
31. Sawaya, R., Schwer, B. & Shuman, S. Structure-function analysis of the yeast NAD+-dependent tRNA 2′-phosphotransferase Tpt1. *RNA* **11**, 107–113 (2005).
32. Stone, P.R. & Hilz, H. Quantitation of hydroxylamine sensitive mono(adenosine diphosphate ribose) residues in different hepatic tissues. *FEBS Lett.* **57**, 209–212 (1975).
33. Barth, H., Aktories, K., Popoff, M.R. & Stiles, B.G. Binary bacterial toxins: biochemistry, biology, and applications of common *Clostridium* and *Bacillus* proteins. *Microbiol. Mol. Biol. Rev.* **68**, 373–402 (2004).
34. Lang, A.E. *et al. Photorhabdus luminescens* toxins ADP-ribosylate actin and RhoA to force actin clustering. *Science* **327**, 1139–1142 (2010).
35. Timinszky, G. *et al.* A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nat. Struct. Mol. Biol.* **16**, 923–929 (2009).
36. Jope, R.S. & Johnson, G.V. The glamour and gloom of glycogen synthase kinase-3. *Trends Biochem. Sci.* **29**, 95–102 (2004).
37. Jackson, M.D. & Denu, J.M. Structural identification of 2′- and 3′-O-acetyl-ADP-ribose as novel metabolites derived from the Sir2 family of β-NAD+-dependent histone/protein deacetylases. *J. Biol. Chem.* **277**, 18535–18544 (2002).
38. Sauve, A.A. *et al.* Chemistry of gene silencing: the mechanism of NAD+-dependent deacetylation reactions. *Biochemistry* **40**, 15456–15463 (2001).
39. Jackson, M.D., Schmidt, M.T., Oppenheimer, N.J. & Denu, J.M. Mechanism of nicotinamide inhibition and transglycosidation by Sir2 histone/protein deacetylases. *J. Biol. Chem.* **278**, 50985–50998 (2003).
40. Kleine, H. *et al.* Dynamic subcellular localization of the mono-ADP-ribosyltransferase ARTD10 and interaction with the ubiquitin receptor p62. *Cell Commun. Signal.* **10**, 28 (2012).

ONLINE METHODS

Purification of recombinant proteins. Macrodomain cDNAs were amplified by PCR, cloned into a pET28-GST vector according to standard protocols, recombinantly expressed in *E. coli* BL21 and purified by using Ni Sepharose High Performance beads (Amersham Biosciences) according to the manufacturer's protocol. GSK3β was purified from SF9 cells[8]. Bound MacroD2 was cleaved from the beads with PreScission protease 3C (GE Healthcare) or eluted with 200 mM imidazole. Protein concentration was determined with a spectrophotometer ND-1000 (Nanodrop), and cleavage was verified by SDS-PAGE. ARTD10 was purified by using the TAP-tagging method as described before[7]. Cloning and purification of mMacro1 and mMacro2 of ARTD8 has been described[41].

De-ADP-ribosylation assay with recombinant proteins. Unless otherwise stated, 50 pmol recombinant purified GST-ARTD10 or His-ARTD1 were incubated with 100 nM $[^{32}\text{P}]$nicotinamide adenine dinucleotide ($[^{32}\text{P}]\text{NAD}^+$, PerkinElmer) for 15 min at 30 °C in reaction buffer (10 mM potassium phosphate, pH 7.2, 10 mM MgCl_2, 1.25 mM DTT, 1 µg/ml pepstatin, 1 µg/ml bestatin, 1 µg/ml leupeptin) and in the case of His-ARTD1 supplemented with 5 pmol annealed double-stranded oligomer (5′-GGAATTCC-3′). The reaction was stopped by filtration through a G50 column (GE Healthcare). De-ADP-ribosylation reactions were performed with 10 pmol MacroD2 protein at 30 °C for 15 min and stopped by the addition of SDS-PAGE loading buffer and boiling (5 min, 95 °C). De-ADP-ribosylation of automodified proteins was visualized by SDS-PAGE and autoradiography. Bands were quantified by using GelEval (http://www.frogdance.dundee.ac.uk).

ADP-ribosylation assays with immunoprecipitated proteins. ADP-ribosylation assays were carried out at 30 °C for 30 min. The reaction mixture (50 mM Tris-HCl, pH 8.0, 0.2 mM DTT, 5 mM MgCl_2 and 50 µM β-NAD^+ (Sigma) and 1 µCi $[^{32}\text{P}]$β-NAD^+ (Amersham Biosciences)) was added to IgG beads with immunoprecipitated HA-ARTD10 or HA-ARTD10-ΔK and optionally 0.5 µg substrate protein in a total reaction volume of 30 µl. Reactions were stopped by adding SDS sample buffer and were subsequently boiled and run on SDS-PAGE. Incorporated radioactivity was analyzed by autoradiography.

ADP-ribosylation of actin by bacterial toxins. *In vitro* ADP-ribosylation of actin was performed as reported[42]. Briefly, 2 µg β/γ actin was incubated with either 100 ng recombinant TccC3hvr or 50 ng CDTa in the presence of 100 nM $[^{32}\text{P}]\text{NAD}^+$, 150 µM cold NAD^+ and reaction buffer (5 mM HEPES, pH 7.5, 0.1 mM CaCl2, 0.5 mM NaAc, 0.1 mM ATP). TccC3hvr and CDT1 reactions were incubated for 30 min at room temperature or 37 °C, respectively.

De-ADP-ribosylation assays with immunoprecipitated proteins. ADP-ribosylation assays were terminated by placing on ice and washing with high-salt buffer (50 mM Tris-HCl, pH 8.0, 0.2 mM DTT, 5 mM MgCl_2, 200 mM NaCl). MacroD2 (500 ng) was added to the beads in 30 µl high-salt buffer. After incubation (30 °C, 20 min), the reaction was stopped by addition of SDS sample buffer and boiling for analysis by SDS-PAGE and autoradiography. For subsequent kinase assays, beads with coupled GST-GSK3β were cooled and washed after incubation with MacroD2.

Chemical de-ADP-ribosylation. For chemical de-ADP-ribosylation with hydroxylamine, automodified ARTD10 was supplemented with 0.8 M hydroxylamine in a 1:1 ratio (v/v). The reactions were incubated at 37 °C for 1 h and subsequently stopped by the addition of SDS loading buffer.

GST-macro pulldown assays with ARTD10 proteins. GST- or His-tagged macrodomains were immobilized on glutathione or Ni Sepharose (Amersham Biosciences) at 4 °C (wild type and mutants) and incubated with automodified ARTD10 (full length; 100 ng) or GST-ARTD10 (818–1025; 50 ng) proteins for 2 h at 4 °C in 1,200 µl of binding buffer (100 mM Tris, pH 7.6, 250 mM NaCl, 125 mM KCl, 50 mM KAc, 1.5% NP-40 (high salt) or 75 mM NaCl, 25 mM KCl, 15 mM KAc, 1% NP-40 (low salt), 10% glycerol and protease inhibitors) and washed 5× with binding buffer (1,200 µl) for 25 min at 4 °C. Bound proteins were dissolved by boiling and were loaded on an SDS-PAGE gel for subsequent autoradiography.

Kinase assays. $[\gamma^{32}\text{P}]$ATP was diluted to 0.16 µCi/µl in 250 µM ATP in 3× kinase assay buffer (5 mM MOPS, pH 7.2, 2.5 mM β-glycerophosphate, 1 mM EGTA, 0.4 mM EDTA, 4 mM MgCl_2, 50 µM DTT and 40 ng/µl BSA). GST-GSK3β (25 ng) or precipitate was incubated in a reaction volume of 25 µl (5 µl 0.16 µCi/µl $[^{32}\text{P}]$ATP solution, 5 µg substrate peptide RRRPASVPPSPSLSRHS(pS)HQRR (Millipore)). After incubating (30 °C, 15 min), the reaction was stopped by placing on ice. Aliquots of 10 µl were spotted on P81 paper in duplicate, washed with 0.5% phosphoric acid and air dried before scintillation counting. Data are presented as mean ± s.d. of at least triplicate measurements from representative experiments. Statistical significance was determined by employing two-tailed Student's *t* test.

LC-MS and HPLC analysis of ADP-ribose. For HPLC analysis, released ADP-ribose or ADP-ribose standards were purified over Microcon Ultracel YM-3 columns and subjected to reversed-phase liquid chromatography on an Accucore C18 2.7 µm, 150 × 2.1 mm ID Column. A water-methanol gradient from 0% to 20% MeOH at a flow rate of 200 µl/min was applied. Free ADP-ribose was monitored by UV absorbance at 260 nm. For LC-MS analysis, 10 µM ADP-ribose standard and samples were analyzed by using hydrophilic interaction chromatography (HILIC) coupled to accurate MS. The chromatographic separation of ADP-ribose was performed on a 0.2 µm × 150 mm BEH amide column, using a linear gradient of acetonitrile to water, 0.5 mM ammonium acetate, pH 9. For MS, we used negative mode with a capillary voltage of 1.2 kV. Data were acquired in MS and MS-MS mode. Extracted ion chromatograms from the MS data were generated by using the monoisotopic mass of ADP-ribose adduct [M-H] 558.064 and a mass window of 10 mDa. For relative quantification of ADP-ribose, MS scans in the elution time range of ADP-ribose were combined, and the ion abundance of mass 558.064 was calculated.

Homology modeling of MacroD2. The MacroD2 homology model (obtained by Modeller[43–46]; **Supplementary Note**) with the lowest discrete optimized protein energy assessment score[47] was selected for minimization and molecular-dynamics refinement. Following 0.5 ns of NVT and subsequent 0.5 ns of NPT equilibration during which the protein heavy atoms and protein Cα atoms were, respectively, positionally restrained, two 100-ns trajectories were generated by using different random seeds (MD runs I and II). Trajectory analyses (**Supplementary Fig. 4a–c**) were performed with the MD-analysis tool WORDOM[48,49].

Docking and molecular-dynamics simulations of ADP-ribose in MacroD2. From the MD simulations of the apo-MacroD2 homology model, a trajectory frame was selected that maximized the solvent-accessible surface area of putative binding-site residues within the mixed α/β macrodomain fold[35], and which maintained similar Asp78 $\chi 1$ and $\chi 2$ angles relative to the corresponding residue of Af1521 (PDB 2BFQ).

Water molecules and ions were removed from this frame. AutoDock Vina[50] was employed to dock the ADP-ribose ligand to the mixed α/β fold. The 20 top-ranking poses were minimized with CHARMM by using the CHARMM27 force field for the protein atoms and the CHARMM general force field[51,52] for ADP-ribose. Upon structural superposition of MacroD2 and Af1521, the minimized pose of ADP-ribose having the lowest r.m.s. deviation value relative to the one in Af1521 (PDB 2BFQ) was immersed in a box of TIP3P water molecules and subjected to ten explicit water MD runs of 10 ns each at 300 K.

For each run, the distance between the center of mass of the putative binding site of MacroD2 and that of ADP-ribose was calculated by using WORDOM[48,49]. Residues forming the putative binding site were those having at least one atom within 5 Å of any ligand atom following equilibration.

A total of 65% of the obtained MD trajectory frames presented a binding-site ligand center-of-mass distance of <6 Å; above this distance, the distal ribose of the ligand rarely re-entered the binding site. Unbinding of ADP-ribose within 10 ns in almost half of the runs is consistent with its high µM inhibition of human MacroD1 activity[28].

Residues participating in hydrogen-bonding to ADP-ribose among trajectory frames in which the ligand remained bound were identified by using WORDOM with a distance cutoff of 4.0 Å between donor (D) and acceptor (A) atoms and a D-H··A angle larger than 130°. Predicted interactions of the distal ribose of ADP-ribose are predominantly with the carboxylate of Asp102 and with protein backbone atoms (**Supplementary Fig. 4d,e**).

MacroD2 residues forming stable hydrogen bonds to structural water molecules were identified by using the GROMACS g_hbond function. The most stable water molecules in the vicinity of the distal ribose are shown (**Supplementary Fig. 4f**) along with their interaction partners in MacroD2.

Figures were created with Pymol (http://pymol.sourceforge.net/).

41. Forst, A.H. *et al.* Recognition of mono-ADP-ribosylated ARTD10 substrates by ART8 macrodomains. *Structure* (in the press).
42. Gülke, I. *et al.* Characterization of the enzymatic component of the ADP-ribosyltransferase toxin CDTa from *Clostridium difficile. Infect. Immun.* **69**, 6004–6011 (2001).
43. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. in *Curr. Protoc. Bioinformatics* **15**, 5.6 (2006).
44. Martí-Renom, M.A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325 (2000).
45. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
46. Fiser, A., Do, R.K. & Sali, A. Modeling of loops in protein structures. *Protein Sci.* **9**, 1753–1773 (2000).
47. Shen, M.Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).
48. Seeber, M., Cecchini, M., Rao, F., Settanni, G. & Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **23**, 2625–2627 (2007).
49. Seeber, M. *et al.* Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* **32**, 1183–1194 (2011).
50. Trott, O. & Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
51. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
52. Brooks, B.R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

doi:10.1038/nsmb.2521

Nat. Struct. Mol. Biol. 20, 502–507 (2013)

Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases

Florian Rosenthal, Karla L H Feijs, Emilie Frugier, Mario Bonalli, Alexandra H Forst, Ralph Imhof, Hans C Winkler, David Fischer, Amedeo Caflisch, Paul O Hassa, Bernhard Lüscher & Michael O Hottiger

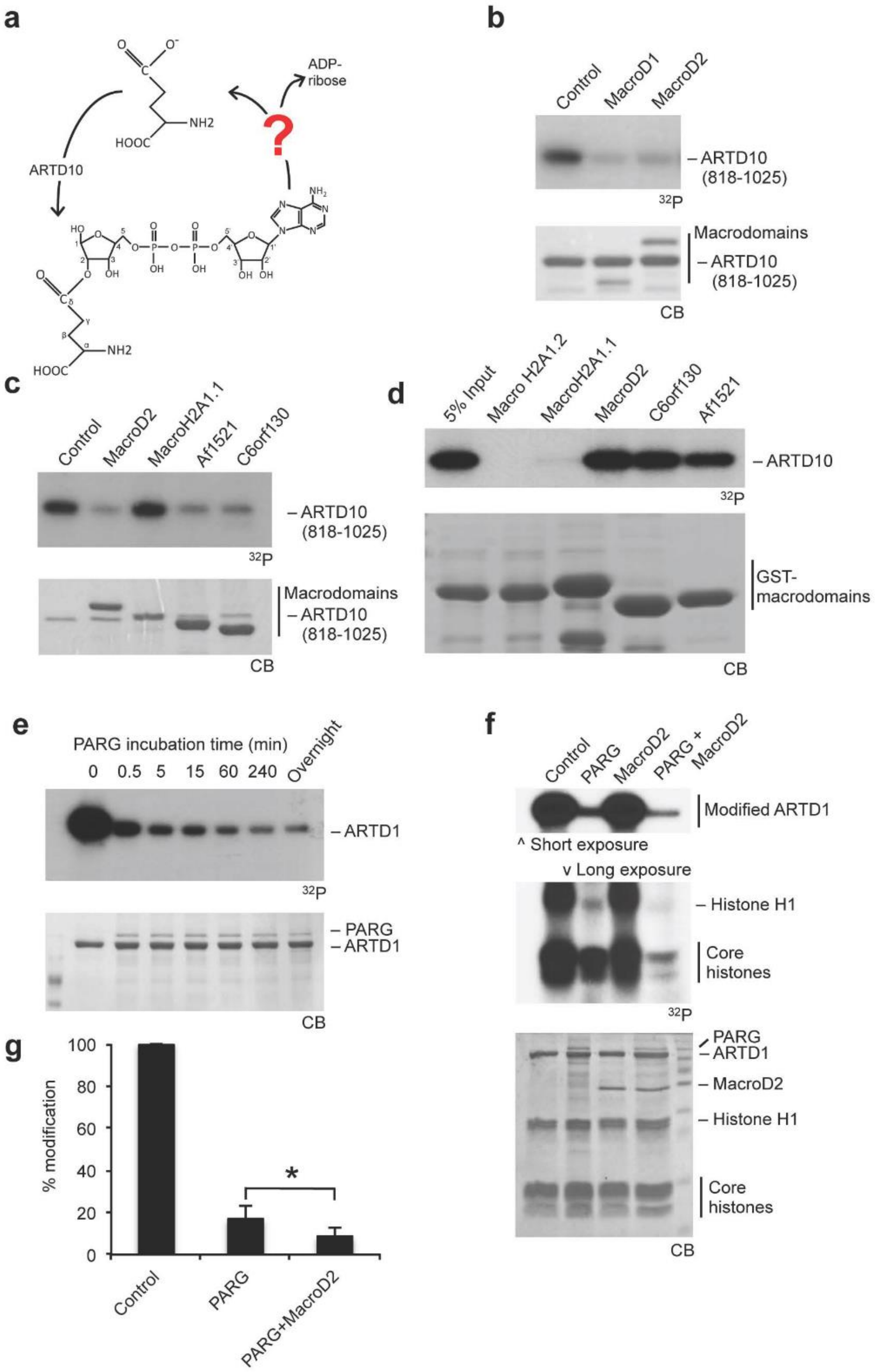
In the version of this supplementary file originally posted online, Ile189 was incorrectly referred to as Ile89 in **Supplementary Figure 4e**. The error has been corrected in this file as of 3 April 2013.

Supplementary material for:

Macrodomain-containing proteins are novel mono-ADP-ribosylhydrolases

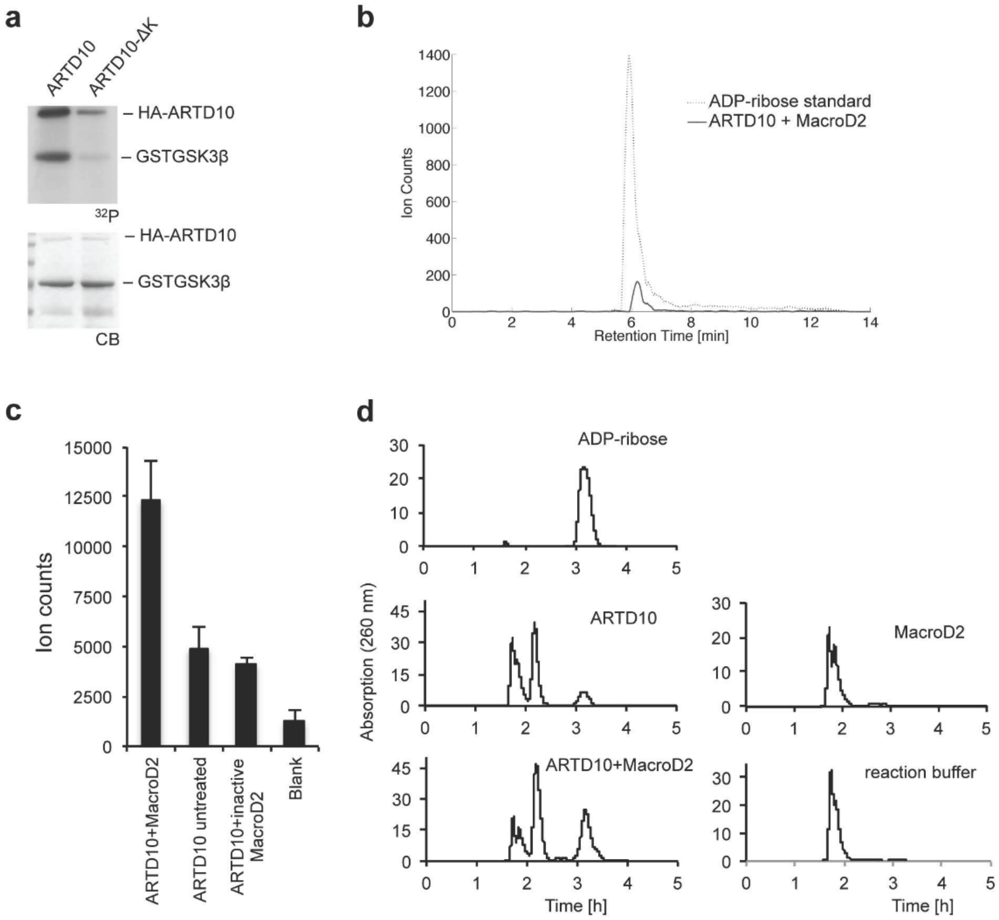
Florian Rosenthal, Karla L.H. Feijs, Emilie Frugier, Mario Bonalli, Alexandra H. Forst, Ralph Imhof, Hans C. Winkler, David Fischer, Amedeo Caflisch, Paul O.Hassa, Bernhard Lüscher and Michael O. Hottiger

Supplementary Figure 1



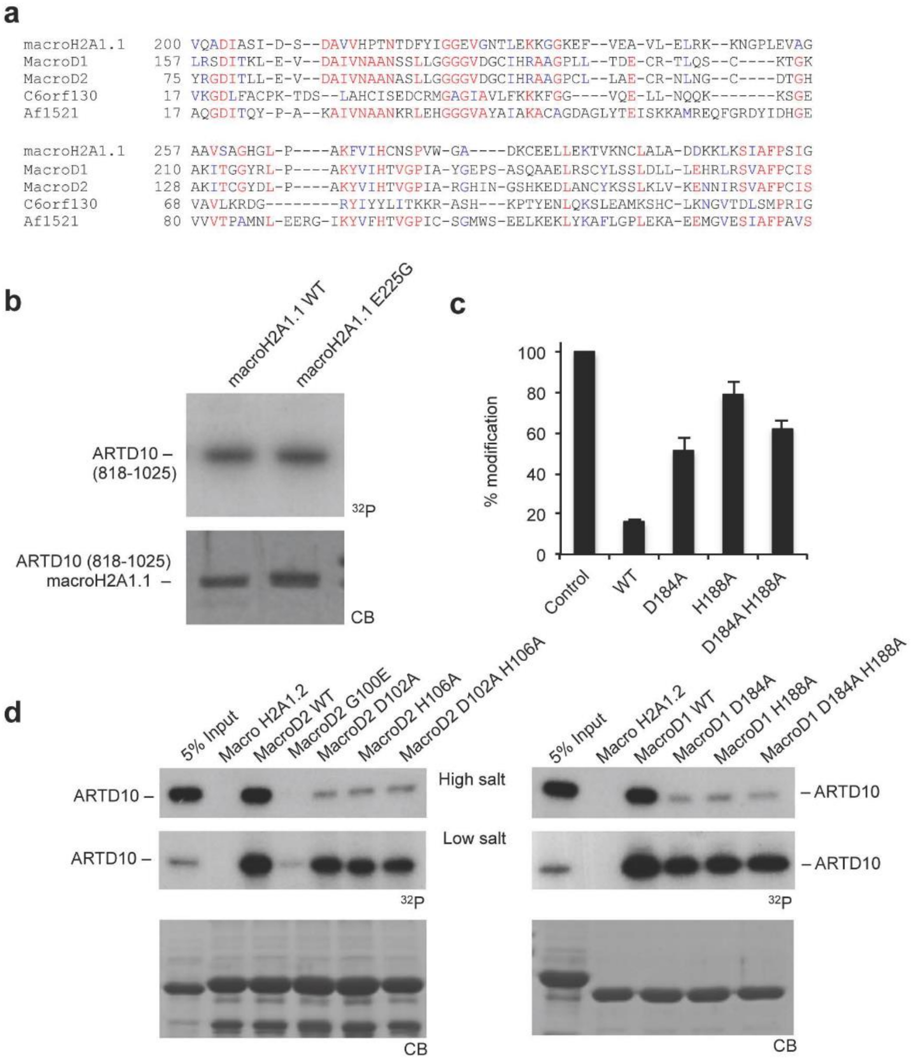
Supplementary Figure 1. Macrodomain-containing proteins hydrolyze mono-ADP-ribose modifications. (a) The missing link in the protein mono-ADP-ribosylglutamate modification cycle. (b-c) MacroD1, MacroD2, Af1521 and C6orf130 hydrolyze mono-ADP-ribosylated catalytic ARTD10 domain (residues 818-1025), while MacroH2A1.1 did not show activity. (d) ADP-ribose binding to different macrodomain proteins. Tagged macrodomain proteins were immobilized and incubated with automodified ARTD10. (e) Protein poly-ADP-ribose hydrolyzing activity of PARG. ARTD1 was auto-ADP-ribosylated *in vitro* with radioactive NAD^+ and subjected to de-ADP-ribosylation assays with recombinant PARG for the indicated times. (f) PARG treatment renders proteins modified by ARTD1 under low NAD^+ concentrations partly susceptible to hydrolysis by MacroD2. (g) MacroD2 further demodifies PARG treated ARTD1. ARTD1 was auto-poly-ADP-ribosylated and consecutively demodified either by PARG alone or by PARG and MacroD2. Three independent experiments were quantified and the results are shown relative to the untreated control (mean ± SD). The effect of MacroD2 treatment on PARG treated samples was significant based on a paired t-test ($p < 0.05$).

Supplementary Figure 2



Supplementary Figure 2. Identification of ADP-ribose as the product released by MacroD2 from auto-modified ARTD10. (a) The ARTD10-ΔK mutant shows reduced ADP-ribosylation activity toward itself and GSK3β. (b) Extracted ion chromatograms of 558.064, monoisotopic mass of released ADP-ribose and ADP-ribose standard [M-H] adduct. Elution time window of ADP-Ribose can be observed between 6 and 7 minutes. (c) Relative amounts of ADP-ribose. MS scans in the elution time range of ADP-ribose were combined and the ion abundance of mass 558.064 is displayed. Error bars: standard deviation of three replicate LC-MS runs. (d) MacroD2 releases ADP-ribose from mono-ADP-ribosylated ARTD10 substrate. Reactions products were separated by HPLC on a RP C18 column and detected by measuring absorption at 260 nm.

Supplementary Figure 3



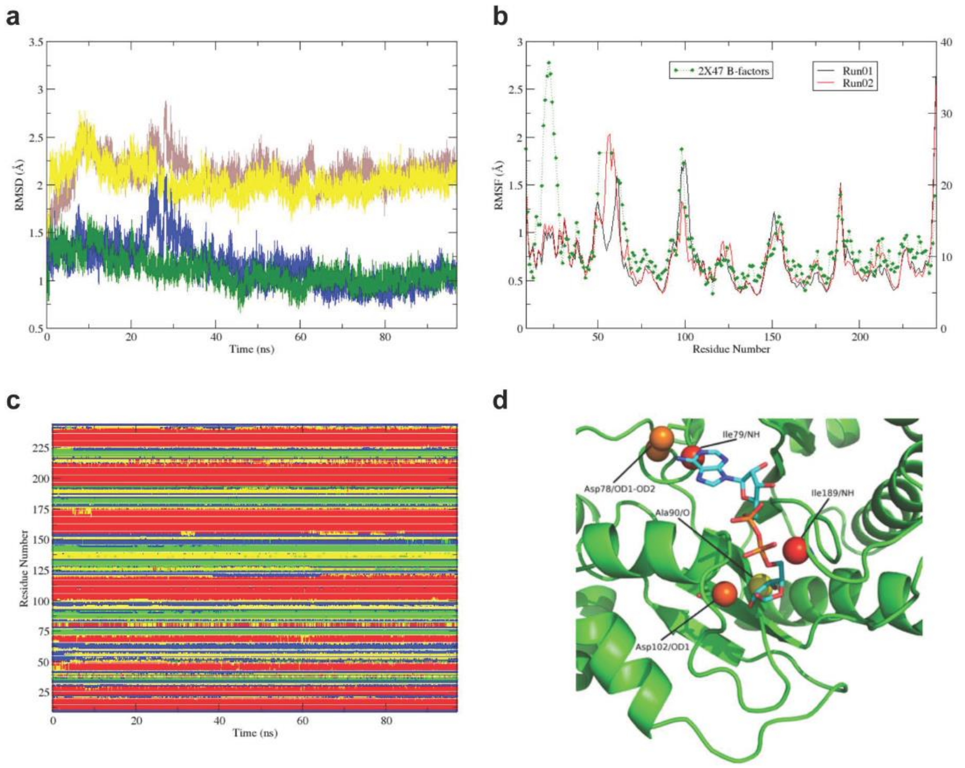
```
macroH2A1.1  200 VQADIASI-D-S--DAVVHPTNTDFYIGGEVGNTLEKKGGKEF--VEA-VL-ELRK--KNGPLEVAG
MacroD1      157 LRSDITKL-E-V--DAIVNAANSSLLGGGGVDGCIHRAAGPLL--TDE-CR-TLQS--C----KTGK
MacroD2       75 YRGDITLL-E-V--DAIVNAANASLLGGGGVDGCIHRAAGPCL--LAE-CR-NLNG--C----DTGH
C6orf130      17 VKGDLFACPK-TDS--LAHCISEDCRMGAGIAVLFKKKFGG----VQE-LL-NQQK-------KSGE
Af1521        17 AQGDITQY-P-A--KAIVNAANKRLEHGGGVAYAIAKACAGDAGLYTEISKKAMREQFGRDYIDHGE

macroH2A1.1  257 AAVSAGHGL-P----AKFVIHCNSPVW-GA----DKCEELLEKTVKNCLALA-DDKKLKSIAFPSIG
MacroD1      210 AKITGGYRL-P----AKYVIHTVGPIA-YGEPS-ASQAAELRSCYLSSLDLL-LEHRLRSVAFPCIS
MacroD2      128 AKITCGYDL-P----AKYVIHTVGPIA-RGHIN-GSHKEDLANCYKSSLKLV-KENNIRSVAFPCIS
C6orf130      68 VAVLKRDG--------RYIYYLITKKR-ASH---KPTYENLQKSLEAMKSHC-LKNGVTDLSMPRIG
Af1521        80 VVVTPAMNL-EERG-IKYVFHTVGPIC-SGMWS-EELKEKLYKAFLGPLEKA-EEMGVESIAFPAVS
```

Supplementary Figure 3. Identification of residues implicated in catalysis.

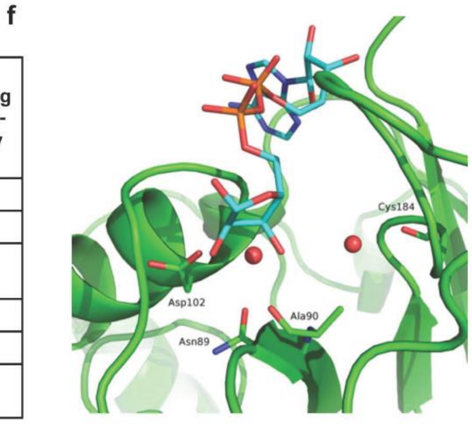
(a) Sequence alignment of macrodomain containing proteins from *Homo sapiens* (macroH2A 1.1, MacroD1, MacroD2, C6orf130), and *Archaeoglobolus fulgidus* (Af1521). Only the amino acids aligning to the residues 157 - 286 of the MacroD1 macrodomain are shown. (b) Mutation of E225 of macroH2A1.1 to a glycine does not restore mono-ADP-ribose hydrolyzing activity. (c) Mutational analysis of MacroD1 highlights residues implicated in enzymatic activity towards mono-ADP-ribosylated ARTD10 (818-1025). (d) ADP-ribose binding (under high and low salt conditions) to WT and mutant MacroD2 (left) or MacroD1 (right) was assessed with His- and GST-tagged, recombinantly expressed macrodomain proteins that were immobilized on glutathione or Nickel sepharose and incubated with automodified ARTD10.

Supplementary Figure 4



e

ADP-ribose atoms	MacroD2 atoms	Frequency of hydrogen bonding over ADP-ribose-bound trajectory frames (%)
Phosphate P1=O	Ile189 NH	79
Adenine N1	Ile79 NH	74
Adenosine ribose C2 OH	Asp102 OD1	62
Adenine NH_2	Asp78 OD1	49
Adenine NH_2	Asp78 OD2	45
Adenosine ribose C2 OH	Ala90 O	28



Supplementary Figure 4. Validation of the MacroD2 homology model and docking and molecular dynamics simulations of ADP-ribose in MacroD2. (a) Time evolution of Cα atoms Root Mean Square Deviation (RMSD) from the MacroD2 homology model. Yellow and brown lines: Cα atoms of residues having an RMSF value of less than 1 Å (shown in panel B) for independent molecular dynamics runs I and II, respectively. Green and blue lines: Cα atoms of the putative binding site for independent molecular dynamics runs I and II, respectively. (b) Cα atom Root Mean Square Fluctuations (RMSF) along the sequence of MacroD2 for molecular dynamics runs I (black) and II (red). B-Factors of template Human MacroD1 (PDB ID: 2X47) are shown in green. (c) Time evolution of secondary structure [1,2] of MacroD2 for molecular dynamics run I. Red: α-helix. Green: β-sheet or β-bridge. Yellow: bend or helix-turn. Blue: loop. The secondary structure evolution was likewise stable for run II. (d) MacroD2 atoms participating in among frames in which ADP-ribose was in the binding site. Atoms are coloured on a red (high)-yellow (low) scale according to frequency of hydrogen bond occurrence. (e) Quantification of frequency of hydrogen-bonding of ADP-ribose to residues of MacroD2. (f) Water molecules close to ADP-ribose forming stable hydrogen-bonding interactions with residues of MacroD2.

Supplementary note:

Homology modeling of MacroD2

Modeling of the three-dimensional structure of residues 9 to 244 of the human MacroD2 protein was undertaken using Modeller (version 9.10) [3-6]. As template, the coordinates of residues 90 to 325 of apo human MacroD1 (PDB ID: 2X47) were selected following alignment of the FASTA sequences of these residues (UniProt ID: Q9BQ69) and residues 9 to 244 of human MacroD2 (UniProt ID: A1Z1Q3) using CLUSTAL O version 1.1.0 on UniProt [7], giving an identity of 53.6%.

The MacroD2 homology model having the lowest Discrete Optimized Protein Energy assessment score [8] was subject to steepest-descent minimisation and explicit-water molecular dynamics (MD) simulations using GROMACS [9] (version 4.5.4) with the CHARMM22 force field [10] and TIP3P water model [11]. Aspartate and glutamate side-chains were negatively charged and those of lysines and arginines were positively charged, while histidines were maintained as neutral. The N-terminus of the protein was acetylated and the C-terminus N-methyl amidated. The protein was immersed in an orthorhombic box of pre-equilibrated water molecules having a minimal distance of 12 Å between the boundary and any protein atom. Sodium and chloride ions were added to the systems for an ion concentration of 150 mM.

The van der Waals and short-range electrostatic interactions were determined up to a cutoff of 10 Å. Periodic boundary conditions were applied and long-range electrostatic interactions were evaluated using the particle-mesh Ewald summation method [12]. MD simulations were performed in the NPT-ensemble at constant temperature (310 K) using the modified Berendsen thermostat, and constant pressure (1 atm) using the Parrinello-Rahman barostat [13]. Covalent bonds involving hydrogen atoms were constrained using the Linear Constraint Solver algorithm [14] and the time-step was 2 fs.

References

1 Andersen, C. A., Palmer, A. G., Brunak, S. & Rost, B. Continuum secondary structure captures protein flexibility. *Structure* **10**, 175-184, (2002).

2 Carter, P., Andersen, C. A. & Rost, B. DSSPcont: Continuous secondary structure assignments for proteins. *Nucleic Acids Res* **31**, 3293-3295, (2003).

3 Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Current Protoc Bioinformatics* **Chapter 5**, Unit 5 6, (2006).

4 Marti-Renom, M. A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **29**, 291-325, (2000).

5 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815, (1993).

6 Fiser, A., Do, R. K. & Sali, A. Modeling of loops in protein structures. *Protein Sci* **9**, 1753-1773, (2000).

7 Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **40**, D71-75, (2012).

8 Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* **15**, 2507-2524, (2006).

9 Hess, B., Kutzner, C., van der Spoel, D. & E., L. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput.* **4**, 435-447, (2008).

10 MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J Physical Chem B* **102**, 3586-3616, (1998).

11 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926-935, (1983).

12 Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys* **103**, 8577, (1995).

13 Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52**, 7182, (1981).

14 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem.* **18**, 1463-1472, (1997).

5. PYRROLO[3,2-B]QUINOXALINE-DERIVATIVES AS TYPE I1/2 AND II EPH TYROSINE KINASE INHIBITORS: STRUCTURE-BASED DESIGN, SYNTHESIS, AND IN-VIVO VALIDATION

Unzue, A., Dong, J., Lafleur, K., Zhao, H., Frugier, E., Caflisch, A. and Nevado C. *J. Med. Chem.,* **2014**, 57, 6834-6844

Journal of Medicinal Chemistry

Article

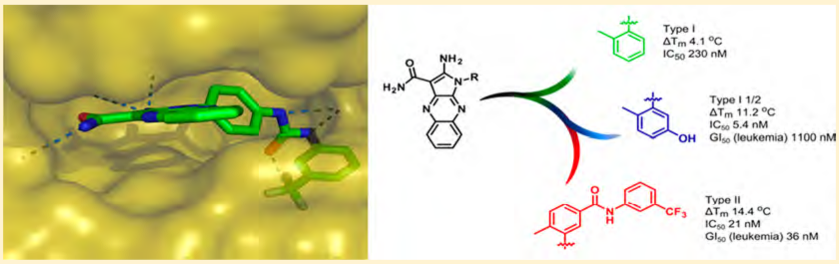
pubs.acs.org/jmc

Pyrrolo[3,2-*b*]quinoxaline Derivatives as Types $I_{1/2}$ and II Eph Tyrosine Kinase Inhibitors: Structure-Based Design, Synthesis, and *in Vivo* Validation

Andrea Unzue,[†,§] Jing Dong,[‡,§] Karine Lafleur,[†] Hongtao Zhao,[‡] Emilie Frugier,[‡] Amedeo Caflisch,*[,‡] and Cristina Nevado*[,†]

[†]Department of Chemistry and [‡]Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Ⓢ *Supporting Information*



ABSTRACT: The X-ray crystal structures of the catalytic domain of the EphA3 tyrosine kinase in complex with two type I inhibitors previously discovered *in silico* (compounds **A** and **B**) were used to design type $I_{1/2}$ and II inhibitors. Chemical synthesis of about 25 derivatives culminated in the discovery of compounds **11d** (type $I_{1/2}$), **7b**, and **7g** (both of type II), which have low-nanomolar affinity for Eph kinases *in vitro* and a good selectivity profile on a panel of 453 human kinases (395 nonmutant). Surface plasmon resonance measurements show a very slow unbinding rate (1/115 min) for inhibitor **7m**. Slow dissociation is consistent with a type II binding mode in which the hydrophobic moiety (trifluoromethyl-benzene) of the inhibitor is deeply buried in a cavity originating from the displacement of the Phe side chain of the so-called DFG motif as observed in the crystal structure of compound **7m**. The inhibitor **11d** displayed good *in vivo* efficacy in a human breast cancer xenograft.

I. INTRODUCTION

Several protein kinases are relevant targets for the treatment of diseases ranging from cancer, inflammation, and cardiovascular conditions to immune related disorders.[1,2] Over the past decade, more than 13 small-molecule kinase inhibitors have been approved by the FDA as therapeutics for various human pathologies.[3−6] In this context, receptor tyrosine kinases play a prominent role, as they are involved in a number of biologically relevant processes for cancer development including oncogenic regulation, cell signal transduction, proliferation, and survival among many others.[7,8] Although irreversible inhibitors that form covalent bonds with cysteine or other nucleophilic residues in the ATP-binding pocket have been recently explored,[9,10] ATP competitive, noncovalent inhibitors are much more abundant and, depending on the binding mode with their protein target, are classified as type I−IV.[11] Most kinase inhibitor drugs are of type I, i.e., they are direct competitors of ATP within the catalytic site of the phosphorylated active conformation of the protein.[12] However, because of the strong similarities between the ATP binding pocket of all human kinases, alternative approaches providing selective binders have been sought.[13−18] Small molecules forming additional interactions with hydrophobic regions adjacent to the ATP binding site are termed type $I_{1/2}$ inhibitors. Alternatively, type II inhibitors target the kinase catalytic site but bind to the inactive conformation of the protein, thus exploring a pocket generated upon displacement of the phenylalanine side chains of the DFG motif.[4] Type III inhibitors, also known as allosteric inhibitors, target areas of the kinase not related to the catalytic domain, whereas type IV do so without competing with ATP. Higher degrees of selectivity are to be expected with the latter two inhibitor types.[19]

Recently, our groups have focused on the *in silico* design, synthesis, and computational-aided optimization of potent and selective receptor tyrosine kinase inhibitors. Successful campaigns have yielded single-digit nanomolar EphB4 inhibitors whose potential antiproliferative activities have been characterized by cellular assays.[20−22] Furthermore, the predicted binding mode could also be confirmed by X-ray diffraction analysis of their complexes with EphA3. Given the

Received: June 18, 2014
Published: July 30, 2014

dx.doi.org/10.1021/jm5009242 | *J. Med. Chem.* 2014, 57, 6834−6844

critical role of Eph receptors and (Eph)−ephrin signaling in tumor growth and progression,[23] a subset of these compounds are currently being pursued toward preclinical development.

Here we describe a multidisciplinary campaign toward the design of novel and potent, type $I_{1/2}$ and II tyrosine kinase inhibitors based on the crystal structure of two type I inhibitors. The parent pyrrolo[3,2-*b*]quinoxaline scaffold was decorated with characteristic functional groups present in previously successful type II binders, thus speeding up the hit to lead optimization campaign. The binding kinetics of the low-nanomolar derivatives **11d** (type I1/2) and **7m** (type II) were characterized by surface plasmon resonance (SPR) measurements. Extensive profiling by biochemical (competition binding) and cellular assays, together with pharmacokinetic measurements in mice resulted in the prioritization of inhibitor **11d** for final validation *in vivo* by a human breast cancer xenograft.

II. CRYSTAL STRUCTURES OF TYPE I INHIBITORS A AND B WITH EPHA3

II.1. Docking Validation by X-ray Diffraction Analysis of Binding Complex. Recently, we reported the discovery of two type I EphB4 inhibitors **A** and **B** by automated docking.[24] The *in silico* predicted binding mode of these molecules is confirmed here by X-ray diffraction analysis of the catalytic domain of EphA3 in complex with both **A** and **B** (Figure 1).

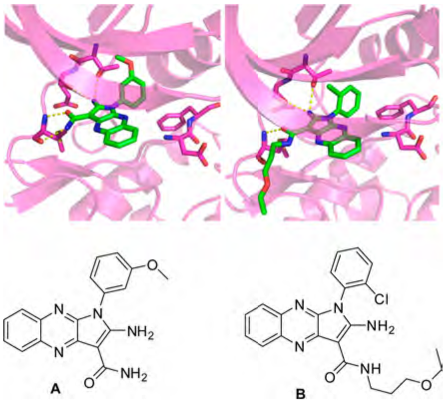


Figure 1. Crystal structures of the catalytic domain of the tyrosine kinase EphA3 in complex with the high-nanomolar inhibitors **A** (left, pdb code 4P4C) and **B** (right, pdb code 4P5Q). The ATP binding site of the EphA3 kinase is shown in magenta ribbons, while the side chains mentioned in the text and the inhibitors are shown by sticks (with carbon atoms in magenta and green, respectively).

The pyrrolo[3,2-*b*]quinoxaline scaffold occupies the ATP binding site with the phenyl substituent nestled into the so-called hydrophobic pocket. The amino substituent at position 2 of the pyrrole ring is involved in a bifurcated hydrogen bond with the side chain hydroxyl of the Thr693 gatekeeper and the backbone carbonyl of Glu694. Furthermore, in the structure with inhibitor **A**, the amide substituent at position 3 of the pyrrole ring is optimally oriented for two hydrogen bonds with the backbone polar groups of Met696 so that **A** forms a total of three hydrogen bonds with the backbone of the hinge region. Only two hydrogen bonds with the same region are observed for inhibitor **B** due to the ethoxy-propyl substitution at the nitrogen of the amide whose trans configuration prevents it from acting as donor to the carbonyl group of Met696. The lack of this hydrogen bond is consistent with the about 10-fold weaker affinity of inhibitor **B** with respect to **A** (IC_{50} of 300 nM for EphB4),[24] which might also originate, at least in part, from the different substituents of the phenyl ring in the hydrophobic pocket, i.e., $-OCH_3$ and −Cl in **A** and **B**, respectively.

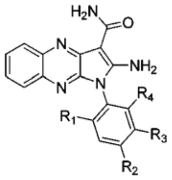
II.2. Design of Type $I_{1/2}$ and Type II Derivatives Based on the X-ray Crystal Structure of the Type I Inhibitors A and B. On the basis of our previous experience[20,22] and earlier reports toward the synthesis of potent type I kinase inhibitors,[25] several modifications within the phenyl ring were designed in order to fine-tune the interactions of the quinoxaline inhibitors with the threonine gatekeeper residue (Thr693) of EphB4. Because of the rather limited space around the phenyl group revealed by the binding modes of **A** and **B**, the introduction of small substituents was envisioned, including the incorporation of a methyl and a hydroxyl group at positions 2 and 5, a combination that had been successfully exploited in our previous studies developing nanomolar inhibitors of EphB4.[20]

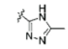
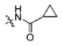
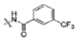
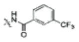
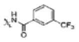
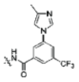
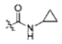
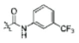
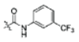
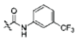
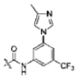
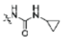
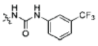
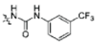
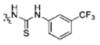
In addition, the binding modes of inhibitors **A** and **B** suggested the possibility to extend the quinoxaline scaffold into the allosteric binding site by substitution of the phenyl ring, resulting in type II inhibitors. Our design campaign targeting the back pocket of the kinase relied on the introduction of a variety of substituents, some of which are present in type II inhibitors. The so-called "tail" moiety of type II inhibitors, located within the allosteric binding site, is characterized by a hydrogen bond donor−acceptor pair (usually an amide or an urea), linked to a hydrophobic substituent that occupies the newly formed pocket in the DFG-out conformation of the kinase.[4,7,12] Amide, urea, and thiourea linkers were incorporated onto the quinoxaline scaffold and attached to a *m*-CF_3-phenyl moiety present in some type II kinase inhibitors, in analogy to AAL993,[26] sorafenib,[27,28] and GNF-5837[29] (**7b**−**d**, **7g**−**i**, **7l**, **7m**, and **12n**). A cyclopropyl ring, a common motif in p38α isoform selective kinase inhibitors,[30−32] was also introduced (**7a**, **7f**, and **7k**). In an effort to increase the hydrophobic interactions within the allosteric binding site, a 4-methyl imidazole ring was added aiming to mimic the well-known drug nilotinib[33] (**7e** and **7j**). In order to increase the solubility, methyl or fluoro substituents were placed in ortho-relative position (**7c**−**e** and **7h**−**j**), therefore distorting the planarity of the molecules.[34] The results in Table 1 underline that, although the identification of substituents successfully binding the allosteric pocket based on known inhibitors was not *a priori* obvious, this was nonetheless an efficient strategy to obtain potent and selective type II kinase inhibitors in a time- and cost-effective manner.

III. CHARACTERIZATION OF NEW TYPE $I_{1/2}$ AND II INHIBITORS

III.1. Synthesis. The synthesis of 1*H*-pyrrolo[2,3-*b*]-quinoxaline derivatives **6** and **7** is shown in Scheme 1. Compound **1** was prepared according to previously reported procedures by condensation of commercially available 2,3-dichloroquinoxaline with malononitrile in the presence of sodium hydride.[35,36] The substitution of the chlorine at position 3 with commercially available anilines **2a**−**i** followed by cyclization afforded intermediates **4a**−**i**.[37] The reaction with synthetically prepared anilines **3a**−**n** delivered tricyclic intermediates **5a**−**n** (Scheme 1).

Table 1. EphA3/EphB4 Inhibition Data for the Synthesized Quinoxaline Derivatives

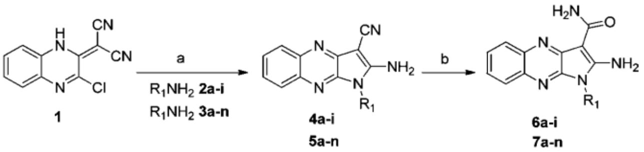


Compound	Type of binding	R_1	R_2	R_3	R_4	Thermal shift (degrees)[a]	FRET[b] enzymatic assay (% of inhibition at 1 μM)	Cellular IC_{50} (nM)[c]
6a	I	Me	H	H	H	4.1	66	230
6b	I	Me	H	H	Me	2.6	36	4400
6c	I	Me	H	H	Cl	2.6	45	2800
6d	I	Me	H	OMe	H	1.5	23	720
6e	$\mathrm{I}_{1/2}$	Me	H	CH_2OH	H	3.6	68	160
6f	I	Cl	H	H	H	3.0	90	-
6g	I	F	H	H	H	2.7	30	-
6h	$\mathrm{I}_{1/2}$	H	H	OH	H	7.2	98	-
6i	$\mathrm{I}_{1/2}$	H	H		H	0.1	7	-
11d	$\mathrm{I}_{1/2}$	Me	H	OH	H	11.2	105	5.4
7a	II	H	H		H	0.3	-	-
7b	II	H	H		H	7.6	-	20
7c	II	Me	H		H	12.3	-	14
7d	II	F	H		H	10.9	-	15
7e	II	Me	H		H	15.8	-	170
7f	II	H	H		H	0.6	-	-
7g	II	H	H		H	9.1	-	20
7h	II	Me	H		H	14.4	-	21
7i	II	F	H		H	13.0	-	24
7j	II	Me	H		H	16.2	-	270
7k	II	H	H		H	0.3	-	-
7l	II	H	H		H	0.7	-	-
7m	II	H		H	H	14.3	-	89
12n	II	H		H	H	9.9	-	560

[a] Average values of triplicate measurements. The standard deviation is smaller than 0.5 degrees. [b] FRET-based enzymatic assay was carried out using the Z-LYTE Kinase Assay Kit—Tyr 1 Peptide (Invitrogen) following the vendor instructions. [c] Cell IC_{50} values were measured in a cellular phosphorylation assay using MEF cells overexpressing EphB4 at Proqinase.

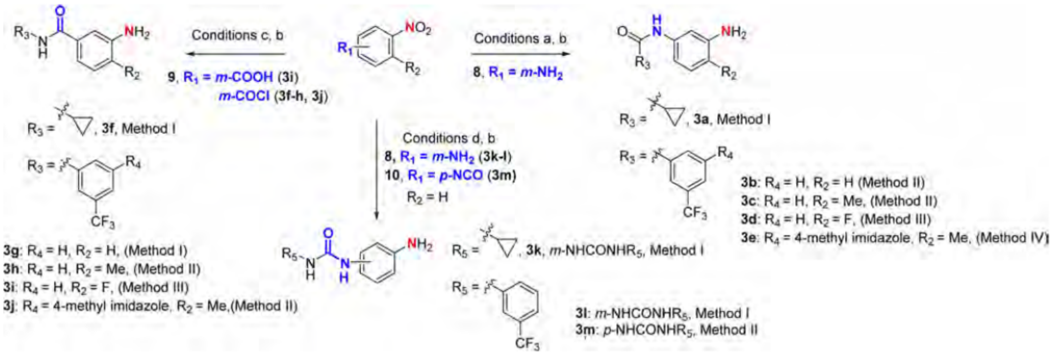
The preparation of the noncommercially available anilines **3a**–**m** used in the above-mentioned condensation reaction has been summarized in Scheme 2. Anilines **3a**–**e**, bearing a N-amide group in meta-relative position, were obtained by

Scheme 1[a]



[a]Reagents and reaction conditions. (a) Protocol I: R_1NH_2 **2a–i**, **3a–d**, **3f–i**, **3k–l**, **3n** (1–4 equiv), EtOH/toluene, 1:1, 130–160 °C, 4–12 h, 30–79%. Protocol II: R_1NH_2 **3e**, **3j**, **3m** (1–1.2 equiv), DMF, 80 °C, 12 h, 21–61%. (b) H_2SO_4, 25 °C, 30 min, 7–99%.

Scheme 2[a]



[a]Reagents and reaction conditions. Conditions (a) Method I: cyclopropanecarboxylic acid, TBTU, DIPEA, DMF, 25 °C, 15 h, 38%. Method II: 3-(trifluoromethyl)benzoyl chloride, Et_3N, DCM, 25 °C, 15 h, 54–99%. Method III: 3-(trifluoromethyl)benzoyl chloride, DIPEA, THF, 25 °C, 12 h, 48%. Method IV: benzoic acid, HOBt, EDC, DMF, 25 °C, 12 h, 64%. Conditions (b) for **3a**, **3b**, **3f**, **3g**, **3k** and **3l**, $SnCl_2.H_2O$, EtOH, 100 °C, 2 h, 53–92%; for **3c–e**, **3h–j**, and **3m**, 10% Pd/C (10 wt %), H_2, MeOH, 25 °C, 4–12 h, 44–99%. Conditions (c) Method I: amine, Et_3N, DCM, 25 °C, 15 h, 50–79%. Method II: aniline, DIPEA, THF, 25 °C, 12 h, 57–78%. Method III: (i) SO_2Cl, DCM, reflux, 3 h; (ii) 3-(trifluoromethyl)aniline, DCM, 25 °C, 12 h, 77%. Conditions (d) Method I: aniline, Et_3N, DCM, 25 °C, 1–3 d, 80–92%. Method II: 3-(trifluoromethyl)aniline, Et_3N, THF, 25 °C, 12 h, 60%.

condensation of 3-nitro-substituted anilines **8** with the corresponding readily available benzoyl chlorides (R_3COCl) or benzoic acids (R_3COOH)[38] using different protocols (Conditions a, Methods I–IV). Reduction of the nitro group with either $SnCl_2$ or Pd/C and H_2 (Conditions b, Supporting Information S12) furnished the desired intermediates. Anilines **3f–j** were prepared by condensation of the 3-nitro-substituted benzoic acid or benzoyl chlorides **9** with the corresponding anilines[39] under reaction conditions c. Reduction of the nitro group with $SnCl_2$ or Pd/C and H_2 delivered the corresponding anilines in 30–77% overall yields. Finally, anilines **3k–m**, bearing urea moieties at meta- and para-relative positions, were prepared by condensation of aniline **8** with the corresponding isocyanate (for **3k–l**) or by condensation of the isocyanate **10** with the respective commercially available aniline under reaction conditions d. Reduction of the nitro group under the conditions described above furnished the corresponding aniline-containing ureas **3k**, **3l**, and **3m** in 42, 85, and 41% yields, respectively. Aniline **3n** was prepared according to a previously reported procedure.[40]

Hydrolysis of the cyano group under strong acidic conditions furnished the desired type I and $I_{1/2}$ inhibitors **6a–i** and type II inhibitors **7a–n**, respectively. The presence of a cyano group in compounds **4** and **5** was confirmed by the presence of a characteristic IR band at around 2200 cm^{-1} and a ^{13}C NMR signal at 115–116 ppm, whereas the amino group appeared as a broad signal at 8 to 8.5 ppm in ^1H NMR.[37]

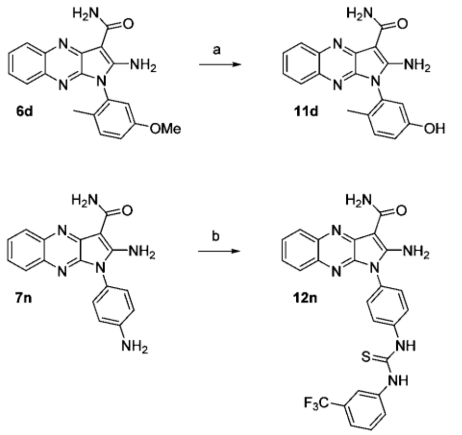
Two more inhibitors were prepared, first by demethylation of **6d** in the presence of BBr_3 under reflux to give the corresponding phenol derivative **11d**, and a second one by condensation of the aniline **7n** with *m*-CF_3-phenyl isothiocyanate, to give the corresponding thiourea **12n** (Scheme 3).

III.2. Biophysical Characterization. The thermodynamics and kinetics of binding of the designed quinoxaline inhibitors were investigated by an array of biophysical techniques including differential scanning fluorimetry, fluorescence resonance energy transfer (FRET) based enzymatic assays, and surface plasmon resonance (SPR). Differential scanning fluorimetry is a high throughput technique in which the increase in thermal stability of a folded protein upon ligand binding is detected by a fluorescent dye while measuring its melting temperature during denaturation.[41]

In order to allow the binding of type II inhibitors to the inactive form of the kinase, the protein (EphA3) was incubated in the presence of the compounds for 1 h. The results (shown in Table 1) highlight the differences in binding between type I, $I_{1/2}$, and II compounds, with type I being the weakest binders (1.5–4.1 °C) and type II the most potent (with up to 16 °C thermal shifts).

For type I inhibitors, the largest thermal shifts (ca. 4 °C) were obtained for *ortho*-methyl (**6a**) and *ortho*-chlorine (**6f**)

Scheme 3[a]



[a] Reagents and conditions: (a) BBr_3, DCM, 130 °C, 4 h, 68%; (b) *m*-CF_3-phenyl isothiocyanate, DMF, 25 °C, 12 h, 69%.

substituted quinoxalines, which is in agreement with previously reported kinase inhibitors[20,42−47] and could be a consequence of restricting the accessible conformations of the phenyl ring as we have previously reported.[20] However, the *ortho*-fluoro substituted inhibitor **6g** or bis-*ortho* substituted **6b** and **6c** caused a lower stabilization of the protein probably due to the small size of fluorine or the introduction of an extra steric bulk, respectively.

The transition from type I to type $I_{1/2}$ by the presence of a hydrogen bond donor at position R_3, and therefore the formation of hydrogen bonds with Glu664 and Asp758, resulted in a remarkable increase in thermal shift for compound **6h** (7.2 °C), which became more pronounced upon introduction of a methyl group in the ortho-relative position (following the trend observed in type I inhibitors) to yield **11d** with a thermal shift of 11.2 °C.

Type II compounds bearing an amide linker followed by a *m*-CF_3-phenyl group caused a similar stabilization effect in the protein as the type $I_{1/2}$ inhibitor **11d**. As expected from previous results with type I and $I_{1/2}$ compounds, the addition of a methyl (**7c** and **7h**) or fluorine (**7d** and **7i**) substituent in R_1 lead to higher thermal shifts. Interestingly, 3-amides (**7g−i**) triggered a higher stabilization of the kinase than 1-amides (**7b−d**), which could indicate the formation of a more favorable hydrogen bond with Glu664. The presence of imidazoles within the allosteric binding site led to compounds **7e** and **7j**, which showed the most promising thermal shifts (16 °C). Urea (**7m**) or thiourea (**12n**) linkers located in para-relative position retained or even enhanced (in the case of urea **7m**) the binding affinity, whereas compound **7l**, bearing the urea in meta-relative position barely presented any thermal shift, suggesting a disruption or a nonfavorable hydrogen bond interaction with Glu664. The replacement of the *m*-CF_3-phenyl group by a cyclopropyl ring proved to be detrimental in all cases, and no thermal shift was observed for products **7a**, **7f**, and **7k**.

The inhibitory activities of type I and type $I_{1/2}$ inhibitors were further evaluated on an enzymatic assay based on fluorescence resonance transfer (FRET) at a single concentration (1 μM, Table 1, column 8). The results were in line with the thermal shifts described above. Compounds **6a** and **6f** showed inhibitory activities higher than 66% at 1 μM (Table 1). However, substitution at R_1 by a smaller fluorine atom yielded compound **6g** with lower binding affinity (30%). Along the same lines, bis-*ortho* substitutions with either methyl or chlorine (**6b,c**) decreased the binding affinity (36 and 45% inhibitory activities, respectively) probably due to unfavorable steric effects. The presence of a hydrogen bond donor at R_3 (type $I_{1/2}$ inhibitors) either as a phenol (**6h** and **11d**) or methylene alcohol (**6e**) greatly improved the inhibitory activity of the molecules thanks to the formation of hydrogen bonds with Glu664 and Asp758 (68−105%). To our surprise, a triazole group at the same position (**6i**) dramatically decreased the binding affinity (7%).

The kinetics of binding of the optimized type $I_{1/2}$ (**11d**) and type II urea inhibitor (**7m**) were investigated using SPR. Upon titrating **11d** and **7m** over immobilized dephosphorylated EphA3, dissociation constants (K_D) in the low nanomolar range were obtained (8.6 and 39.3 nM, respectively), confirming the high affinity of the compounds (Figure 2). One of the advantageous characteristics of type II inhibitors over type I or $I_{1/2}$ is their slow dissociation rate from the target protein,[48−50] as demonstrated by the remarkably slow k_{off} measured for **7m** (1.45×10^{-4} s^{-1}) in comparison to the type $I_{1/2}$ **11d** (1.52×10^{-3} s^{-1}). The slow k_{off} of **7m** corresponds to a residence time of 115 min, a value that compares positively with that of marketed drugs such as imatinib (28 min for dephosphorylated ABL), nilotinib (202 and 205 min for dephosphorylated and phosphorylated ABL, respectively), and dasatinib (15 and 4 min).[51] The long residence time of type II inhibitors is considered to be beneficial for drug efficacy and selectivity *in vivo* due to the high concentration of the drug near the

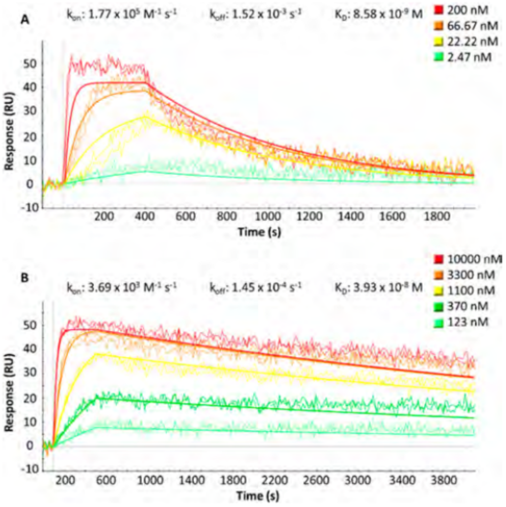


Figure 2. SPR analysis of the binding of **11d** (A) and **7m** (B) to EphA3. A 3-fold serial dilution of the compounds was made starting from 200 nM (for **11d**) and 10 μM (for **7m**) in duplicate. The theoretical global fit to a 1:1 interaction model is shown in straight lines.

target,[50,52] as described for the EGFR-specific inhibitor lapatinib.[53]

IV. VALIDATION OF TYPE II BINDING BY X-RAY CRYSTAL STRUCTURE DETERMINATION

The crystal structure of the catalytic domain of the EphA3 kinase in complex with inhibitor **7m** (solved at 2.0 Å resolution, Supporting Information S34) confirms a type II binding mode, i.e., with the DFG-out conformation (Figure 3). The pyrrolo[3,2-*b*]quinoxaline scaffold occupies the ATP binding site and is involved in the same hydrogen bonds with the hinge region as the type I inhibitor **A** (Figure 1). In addition, the urea linker of inhibitor **7m** acts as hydrogen bond acceptor from the Ser757 side chain and the amide backbone of Asp758, and hydrogen bond donor to the side chain of Glu664. The *m*-CF_3-phenyl moiety is nestled in the hydrophobic pocket, which originates from the displacement of the Phe side chain of the DFG motif. Thus, the type II binding mode of compound **7m** validates our design based on the crystal structures of the complexes with the type I inhibitors (*vide supra*, section II.2).

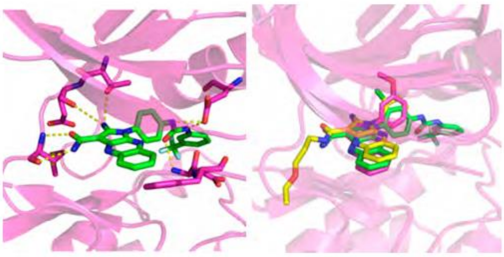


Figure 3. Crystal structures of the catalytic domain of the tyrosine kinase EphA3 in complex with the low-nanomolar inhibitor **7m** (left, pdb code 4P5Z) and superposition of the three inhibitors **A**, **B**, and **7m** based on structural alignment of the C_α atoms of the EphA3 kinase domain (right). The ATP binding site of the EphA3 kinase is shown in magenta ribbons, while the side chains mentioned in the text and the inhibitors are shown by sticks.

V. SELECTIVITY AND CELLULAR ACTIVITY

V.1. Selectivity Profiles from Biochemical Assays. The selectivity profile of inhibitors **11d**, **7b**, and **7g** was determined by an *in vitro* competition binding assay using recombinant kinases (KINOMEscan at DiscoveRx).[54] It is important to note that this assay reports on binding affinity and does not require ATP. The selectivity panel consisted of 453 human kinases, 58 of which were disease related mutant kinases (mainly of ABL1, EGFR, and PIK3CA). Single dose measurements were carried out at 1 μM concentration of the inhibitor.

Compounds **11d**, **7b**, and **7g** present a very similar selectivity profile (Figure 4); in particular, strong binding is only observed for tyrosine kinases with threonine as a gatekeeper residue, e.g., ABL1/2, BRAF, DDR1, EphA/B (all but EphA7, which has a Ile gatekeeper), KIT, LCK, SRC, and YES. The latter data suggest that most (or even all) tyrosine kinases with a Thr gatekeeper can assume the DFG-out conformation. Quantita-

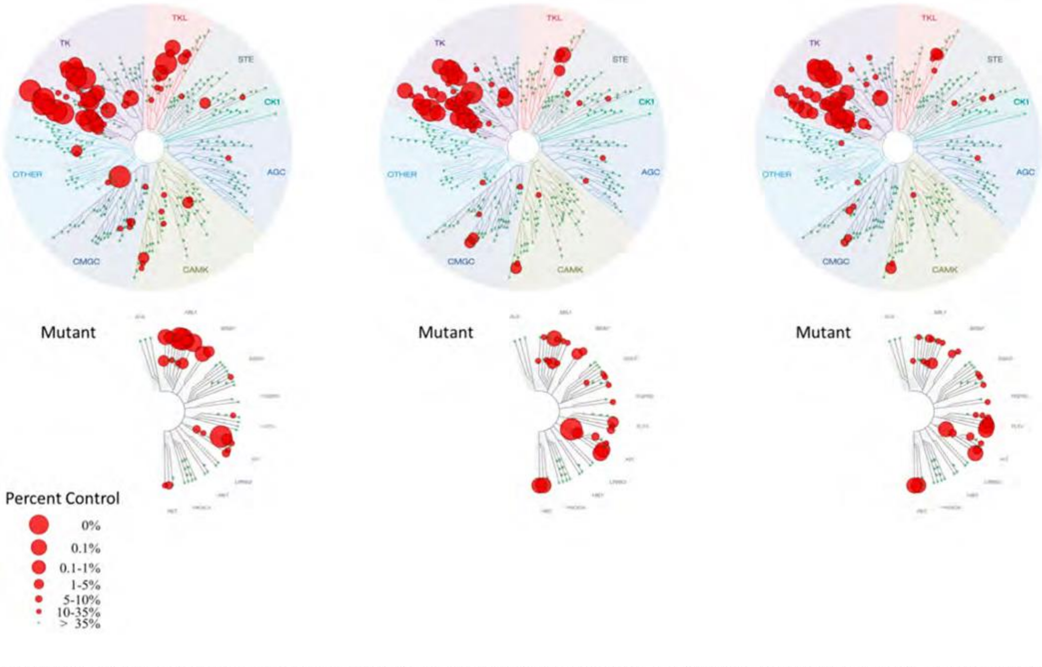


Figure 4. Selectivity profiles of compound **11d** (left), **7b** (center), and **7g** (right) tested on a panel of 395 nonmutant (top) and 58 mutant (bottom) kinases at DiscoveRx. Measurements were performed at a concentration of 1 μM of the inhibitor. The affinity is defined with respect to a DMSO control. The dendrogram was obtained from KinomeScan using the KinomeTree software.

tively, each of the three inhibitors **11d**, **7b**, and **7g** binds with an affinity 10-fold (100-fold) higher than the DMSO negative control to only about 10% (5%) of the 395 wild-type kinases tested. Interestingly, the selectivity profiles of the type $I_{1/2}$ (**11d**) and II (**7b** and **7g**) quinoxaline-based inhibitors is very similar to the one of our previously reported type I and $I_{1/2}$ xanthine-based inhibitors (compounds **40** and **3**, respectively, in ref 22), which is due, at least in part, to the use of an Eph tyrosine kinase (EphB4) as primary target for the *in silico* screening and optimization.

V.2. Cellular Assays. The most potent inhibitors obtained in the optimization campaign were further tested in cell-based assays. Cellular phosphorylation assays on MEF cells transfected with myc-tagged human EphB4 revealed a comparable tendency to the one observed in the enzymatic assay (Table 1, column 9). The type I inhibitors (**6a–d**) displayed cellular IC_{50} values in the 230–4400 nM range, with the *ortho*-methyl substituted derivative **6a** as the most potent member of this series. The type $I_{1/2}$ inhibitor **11d** and type II compounds bearing amide linkers and a *m*-CF_3-phenyl group (**7b–d** and **7g–i**) displayed levels of inhibitory activity in the low nanomolar range (6–24 nM), thus being the most promising molecules of the optimization campaign. In agreement with thermal shift experiments, the presence of a urea (**7m**) and specially the thiourea linker (**12n**) decreased the potency of the compounds (89 and 560 nM, respectively). Interestingly, the imidazole substituted compounds (**7e** and **7j**) proved to be the weakest type II inhibitors (170 and 270 nM, respectively) in contrast to the high thermal shifts obtained (16 °C) in the differential scanning fluorimetry measurements, pointing toward potential cell permeability or efflux issues.

EphB4 overexpression has been linked to several types of cancer, including breast,[55] colon, [56] and ovarian.[57] Compounds **11d** and **7m** were screened against the NCI-60 cancer cell line panel (Supporting Information S3 and S4) displaying antiproliferative activities against leukemia (K-562), lung (HOP-92), colon (HT-29), renal (A498), and breast cancer cells (MDA-MB-231 and HS 578T) in the low nanomolar range. Driven by these results, the most promising inhibitors of our optimization campaign were tested in-house against the above-mentioned NCI cancer cell lines (Table 2). The leukemia K-562 cell line was particularly sensitive toward the optimized type II quinoxaline inhibitors, especially in the case of 3-amide compounds **7h** and **7i**, which showed remarkably low GI_{50} values (36 and 81 nM, respectively). Interestingly, similar levels of potency were found for imidazole substituted compounds **7e** and **7j**, which seemed to be among the weakest type II binders in the cellular phosphorylation assays, possibly indicating other targets than Eph for these molecules. In addition, the potential of **11d** (the most potent compound on cellular phosphorylation assays with an IC_{50} of 6 nM) to inhibit the growth of patient-derived tumor cell lines was studied using a propidium iodide-based proliferation assay and dasatinib as a reference (Oncotest, Table 3). Cell lines included colon, lung, kidney, pancreatic, prostate, and stomach cancer cells. Whereas dasatinib presented double-digit nanomolar activities against RXF 393NL, LXFA 983L, and PRXF DU145, **11d** exhibited low micromolar GI_{50} values, with RXF 393NL being the most sensitive cell line.

Table 2. Antiproliferative Activity against NCI Tumor Cell Lines[a]

compd	MDA -MB-231	K-562	A498	HT29	KM12	HOP-92
6d			59.3	50.3		
6e		44.9	36.4	90.5		
6i			50.9	64.5		
11d	2.64	1.05	5.88	4.59	1.55	0.49
7b	3.09	1.52	13.4	29.4	3.98 (4.03)	4.65 (3.91)
7c	3.88	0.73	2.50	5.87	0.80	4.23
7d	1.44	0.37	1.88	1.80 (9.43)	2.16	2.06
7e	1.93	0.030	4.01	13.4	2.54	5.12
7g	10.8	0.820	10.6	23.1	2.78	5.13
7h	1.32	0.036	2.07	2.82	1.57	1.92
7i	2.69	0.081	2.44	2.97	1.87	2.12
7j	3.05	0.029	5.71	4.36	2.84	3.01
7m	10.3	5.45	10.8	18.0	0.67	10.0
12n	9.63	5.07	8.73	16.3	2.52	14.25

[a] GI_{50} values were determined using resazurin reduction after 2–3 days of incubation with the corresponding compound. GI_{50} values are given in micromolar concentrations (μM) as the mean of at least three independent experiments. Variability around the mean value was <50% unless otherwise indicated by an SE value in parentheses.

Table 3. Antiproliferative Activity against Patient Derived Tumor Cell Lines[a]

compd	**11d**	dasatinib
RXF 393NL	0.725	0.0217
CXF 1103L	3.83	4.36
LXFA 983L	2.22	0.0565
GXF 251L	8.01	2.25
PAXF 1657L	2.92	0.121
PRXF DU145	2.92	0.0623

[a] GI_{50} values were determined at Oncotest using a modified propidium iodide assay. Measurements were performed after 4 days of incubation with **11d** and dasatinib. GI_{50} values are given in micromolar concentrations (μM).

The implication of EphB4-ephrinB2 signaling in sprouting angiogenesis and blood vessel maturation[58] and the inhibition of vascular endothelial growth factor (VEGFR)-driven angiogenesis by the selective EphB4 inhibitor NVP-BHG712,[59] led us to examine the efficacy of **11d** on human endothelial cell sprouting in a spheroid based cellular angiogenesis assay (ProQinase, Supporting Information S36).[60] Compound **11d** was able to successfully inhibit VEGF-A induced HUVEC (primary human umbilical vein endothelial cells) sprouting in a dose dependent manner with an IC_{50} value of 1.5 μM.

VI. *IN VIVO* DATA

Three of the most promising compounds from these series (**7b**, **7g**, and **11d**) were selected for evaluation of pharmacokinetic properties in 20–30 g male CD-1 (ICR) mice on intravenous (IV) and oral (PO) administration. Low to moderate oral bioavailability of tested compounds in mice was observed, with compounds **11d** and **7g** giving the highest values (Table 4). Promising cellular efficacy and pharmacokinetic properties incited the subsequent evaluation of compound **11d** in a xenograft mouse model with a tumor derived from the MDA-MB-231 cell line. High compound clearance (Cl) and moderate half-life ($t_{1/2}$) values determined in the pharmacokinetic study motivated a twice-daily dosing regime totaling 100 mg/kg/day of compound **11d** over 21 days. Median tumor volume

Table 4. Pharmacokinetic Properties in Mice

	11d		7b		7g	
compd	iv	po	iv	po	iv	po
dose (mg/kg)	1	5	1	5	1	5
Cl (mL/min/kg)	42		32		31	
V_{ss} (L/kg)	1.6		2.2		2.2	
$t_{1/2}$ (h)	1.7	1.7	1.2	5.0	1.1	2.8
AUC_{last} (h·ng/mL)	392	493	506	263	533	803
F (%)		25		10		30

progression over time, starting from 108 mm^3, of both treatment and control cohorts is given in Figure 5.

In this study, median treatment-group tumor volume remained essentially stable throughout the treatment period, achieving a median tumor volume of 126 mm^3 at day 19 as opposed to the control group whose median tumor reached 650 mm^3 in the same period. Tumor growth inhibition (%TGI) was statistically significant (Mann−Whitney $U = 0$, $P \leq 0.001$, two-tailed) and quantified at 81% relative to the control group. Mean body weight of the treatment cohort decreased up to 16.3% of the initial mean body weight of this cohort during the treatment period. Treatment with compound **11d** provides a significant limitation in tumor progression over the control, suggesting that further studies of such xenograft model at lower doses of compound **11d** might provide tumor volume control with lessened weight loss. The further evaluation of compounds **11d** on mouse models of K-562 leukemia is underway.

VII. CONCLUSIONS

The X-ray crystal structures of the EphA3 kinase in complex with two high-nanomolar inhibitors based on the 2-amino-1-phenyl-pyrrolo[3,2-*b*]quinoxaline-3-carboxamide scaffold confirmed the type I binding mode obtained previously by automatic docking (Figure 1). This structural information was used to design type $I_{1/2}$ and type II derivatives by taking advantage of the existing knowledge on privileged chemical motifs, i.e., hydroxyl group in meta position of the phenyl ring (for type $I_{1/2}$) and hydrophobic moieties connected to the phenyl ring by amide or urea linkers (on type II). Chemical synthesis of ca. 25 derivatives (Table 1) culminated in several low nanomolar inhibitors with a good selectivity profile (Figure 4). The X-ray crystal structure of the EphA3 kinase in the complex with the inhibitor **7m** (Figure 3) provided the final validation of the structure-based design; in particular, the DFG-out conformation confirmed the type II binding. Moreover, the slow kinetics of unbinding of compound **7m** (measured by SPR, Figure 2) is congruent with the type II binding mode. Three interesting observations emerge from this study. First, it is possible to "elongate" a type $I_{1/2}$ into a type II inhibitor by introducing an amide or urea linked to a bulky hydrophobic group. These type II linkers are involved in the same hydrogen bonds as the type $I_{1/2}$ bearing a hydroxyl group in the same position, while the hydrophobic moiety occupies the pocket resulting from the displacement of the Phe side chain of the DFG motif. The similar selectivity profiles of type $I_{1/2}$ and type II inhibitors indicate that mainly the moiety in contact with the gatekeeper's side chain and hinge region determines specificity. Finally, *in vivo* assays (mice xenografted with human breast cancer) confirmed the cytostatic activity of one of our inhibitors (**11d**), which makes this type $I_{1/2}$ compound a candidate lead for further preclinical development.

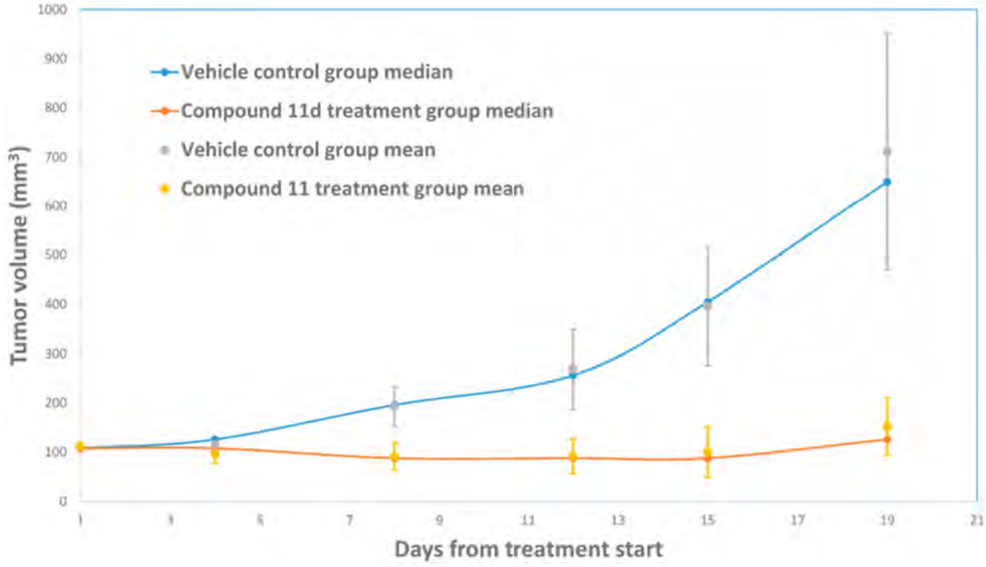


Figure 5. *In vivo* antitumor activity of compound **11d** in MDA-MB-231 nude mice xenografts. The mice received by gavage twice-daily 50 mg/kg of compound **11d** (red) or vehicle control (blue). Each data point is the median or mean of a cohort of 9 animals. Error bars show standard deviations of the mean.

■ ASSOCIATED CONTENT

Supporting Information

General procedures for the synthesis and characterization, biological evaluation of all reported compounds, and X-ray crystal structure refinement data. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

PDB codes for EphA3 in complex with the inhibitors **A**, **B**, and **7m** are 4P4C, 4P5Q, and 4P5Z, respectively.

■ AUTHOR INFORMATION

Corresponding Authors

*(C.N.) E-mail: cristina.nevado@chem.uzh.ch. Phone: (41) 446353945. Fax: (41) 446353948.
*(A.C.) E-mail: caflisch@bioc.uzh.ch.

Author Contributions

§A.U. and J.D. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Danzhi Huang for interesting discussions. We also thank Prof. Sirano Dhe-Paganon for providing the plasmid used for EphA3 expression. The authors would like to thank the Swiss National Science Foundation and the Swiss Cancer League (Krebsliga) for financial support.

■ ABBREVIATIONS USED

Eph, erythropoietin-producing human hepatocellular carcinoma receptor; DFG, aspartate-phenylalanine-glycine; ATP, adenosine triphosphate; SPR, surface plasmon resonance; DMF, dimethylformamide; TBTU, *N,N,N′,N′*-tetramethyl-*O*-(benzotriazol-1-yl)uronium tetrafluoroborate; DIPEA, diisoproylethylamine; DCM, dichloromethane; THF, tetrahydrofuran; HOBt, hydroxybenzotriazole; EDC, *N*-(3-(dimethylamino)propyl)-*N′*-ethylcarbodiimide; FRET, fluorescence-resonance energy transfer; Abl, abelson murine leukemia viral oncogene homologue; EGFR, epidermal growth factor receptor; DDR, discoidin domain receptor; Lck, lymphocyte-specific kinase; DMSO, dimethyl sulfoxide; MEF, mouse embryonic fibroblasts; VEGF, vascular endothelial growth factor; HUVEC, human umbilical vein endothelial cells

■ REFERENCES

(1) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912−1934.

(2) Cohen, P. Protein kinases: the major drug targets of the twenty-first century? *Nat. Rev. Drug Discovery* **2002**, *1*, 309−315.

(3) Li, R.; Stafford, J. A. *Kinase Inhibitor Drugs*; John Wiley & Sons, Inc.: Hoboken, NJ, 2009.

(4) Blanc, J.; Geney, R.; Menet, C. Type II kinase inhibitors: an opportunity in cancer for rational design. *Anticancer Agents Med. Chem.* **2013**, *13*, 731−747.

(5) Kontzias, A.; Kotlyar, A.; Laurence, A.; Changelian, P.; O'Shea, J. J. Jakinibs: a new class of kinase inhibitors in cancer and autoimmune disease. *Curr. Opin. Pharmacol.* **2012**, *12*, 464−470.

(6) O'Brien, Z.; Moghaddam, M. F. Small molecule kinase inhibitors approved by the FDA from 2000 to 2011: a systematic review of preclinical ADME data. *Expert Opin. Drug Metab. Toxicol.* **2013**, *9*, 1597−1612.

(7) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28−39.

(8) Janne, P. A.; Gray, N.; Settleman, J. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat. Rev. Drug Discovery* **2009**, *8*, 709−723.

(9) Liu, Q. S.; Sabnis, Y.; Zhao, Z.; Zhang, T. H.; Buhrlage, S. J.; Jones, L. H.; Gray, N. S. Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* **2013**, *20*, 146−159.

(10) Garuti, L.; Roberti, M.; Bottegoni, G. Irreversible protein kinase inhibitors. *Curr. Med. Chem.* **2011**, *18*, 2981−2994.

(11) Cozza, G.; Bortolato, A.; Menta, E.; Cavalletti, E.; Spinelli, S.; Moro, S. ATP non-competitive Ser/Thr kinase inhibitors as potential anticancer agents. *Anticancer Agents Med. Chem.* **2009**, *9*, 778−786.

(12) Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358−364.

(13) Traxler, P.; Furet, P. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol. Ther.* **1999**, *82*, 195−206.

(14) Huang, D.; Zhou, T.; Lafleur, K.; Nevado, C.; Caflisch, A. Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis. *Bioinformatics* **2010**, *26*, 198−204.

(15) Bamborough, P.; Drewry, D.; Harper, G.; Smith, G. K.; Schneider, K. Assessment of chemical coverage of kinome space and its implications for kinase drug discovery. *J. Med. Chem.* **2008**, *51*, 7898−7914.

(16) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127−132.

(17) Morphy, R. Selectively nonselective kinase inhibition: striking the right balance. *J. Med. Chem.* **2010**, *53*, 1413−1437.

(18) Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011**, *29*, 1046−1051.

(19) Eglen, R.; Reisine, T. Drug discovery and the human kinome: recent trends. *Pharmacol. Ther.* **2011**, *130*, 144−156.

(20) Lafleur, K.; Huang, D.; Zhou, T.; Caflisch, A.; Nevado, C. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J. Med. Chem.* **2009**, *52*, 6433−6446.

(21) Zhao, H. T.; Dong, J.; Lafleur, K.; Nevado, C.; Caflisch, A. Discovery of a novel chemotype of tyrosine kinase inhibitors by fragment-based docking and molecular dynamics. *ACS Med. Chem. Lett.* **2012**, *3*, 834−838.

(22) Lafleur, K.; Dong, J.; Huang, D.; Caflisch, A.; Nevado, C. Optimization of inhibitors of the tyrosine kinase EphB4. 2. Cellular potency improvement and binding mode validation by X-ray crystallography. *J. Med. Chem.* **2013**, *56*, 84−96.

(23) Boyd, A. W.; Bartlett, P. F.; Lackmann, M. Therapeutic targeting of EPH receptors and their ligands. *Nat. Rev. Drug Discovery* **2014**, *13*, 39−62.

(24) Zhao, H.; Huang, D. Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011**, *6*, e19923.

(25) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "gatekeeper door": exploiting the active kinase conformation. *J. Med. Chem.* **2010**, *53*, 2681−2694.

(26) Bold, G.; Altmann, K. H.; Frei, J.; Lang, M.; Manley, P. W.; Traxler, P.; Wietfeld, B.; Bruggen, J.; Buchdunger, E.; Cozens, R.; Ferrari, S.; Furet, P.; Hofmann, F.; Martiny-Baron, G.; Mestan, J.; Rosel, J.; Sills, M.; Stover, D.; Acemoglu, F.; Boss, E.; Emmenegger, R.; Lasser, L.; Masso, E.; Roth, R.; Schlachter, C.; Vetterli, W. New anilinophthalazines as potent and orally well absorbed inhibitors of the VEGF receptor tyrosine kinases useful as antagonists of tumor-driven angiogenesis. *J. Med. Chem.* **2000**, *43*, 2310−2323.

(27) Smith, R. A.; Barbosa, J.; Blum, C. L.; Bobko, M. A.; Caringal, Y. V.; Dally, R.; Johnson, J. S.; Katz, M. E.; Kennure, N.; Kingery-Wood, J.; Lee, W.; Lowinger, T. B.; Lyons, J.; Marsh, V.; Rogers, D. H.; Swartz, S.; Walling, T.; Wild, H. Discovery of heterocyclic ureas as a

new class of raf kinase inhibitors: identification of a second generation lead by a combinatorial chemistry approach. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2775−2778.

(28) Khire, U. R.; Bankston, D.; Barbosa, J.; Brittelli, D. R.; Caringal, Y.; Carlson, R.; Dumas, J.; Gane, T.; Heald, S. L.; Hibner, B.; Johnson, J. S.; Katz, M. E.; Kennure, N.; Kingery-Wood, J.; Lee, W.; Liu, X. G.; Lowinger, T. B.; McAlexander, I.; Monahan, M. K.; Natero, R.; Renick, J.; Riedl, B.; Rong, H.; Sibley, R. N.; Smith, R. A.; Wolanin, D. Omega-carboxypyridyl substituted ureas as Raf kinase inhibitors: SAR of the amide substituent. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 783−786.

(29) Albaugh, P.; Fan, Y.; Mi, Y.; Sun, F. X.; Adrian, F.; Li, N. X.; Jia, Y.; Sarkisova, Y.; Kreusch, A.; Hood, T.; Lu, M.; Liu, G. X.; Huang, S. L.; Liu, Z. S.; Loren, J.; Tuntland, T.; Karanewsky, D. S.; Seidel, H. M.; Molteni, V. Discovery of GNF-5837, a selective TRK inhibitor with efficacy in rodent cancer tumor models. *ACS Med. Chem. Lett.* **2012**, *3*, 140−145.

(30) Liu, C.; Lin, J.; Wrobleski, S. T.; Lin, S.; Hynes, J.; Wu, H.; Dyckman, A. J.; Li, T.; Wityak, J.; Gillooly, K. M.; Pitt, S.; Shen, D. R.; Zhang, R. F.; McIntyre, K. W.; Salter-Cid, L.; Shuster, D. J.; Zhang, H.; Marathe, P. H.; Doweyko, A. M.; Sack, J. S.; Kiefer, S. E.; Kish, K. F.; Newitt, J. A.; McKinnon, M.; Dodd, J. H.; Barrish, J. C.; Schieven, G. L.; Leftheris, K. Discovery of 4-(5-(cyclopropylcarbamoyl)-2-methylphenylamino)-5-methyl-*N*-propylpyrrolo[1,2-f][1,2,4]triazine-6-carboxamide (BMS-582949), a clinical p38α MAP kinase inhibitor for the treatment of inflammatory diseases. *J. Med. Chem.* **2010**, *53*, 6629−6639.

(31) Hynes, J., Jr.; Wu, H.; Pitt, S.; Shen, D. R.; Zhang, R.; Schieven, G. L.; Gillooly, K. M.; Shuster, D. J.; Taylor, T. L.; Yang, X.; McIntyre, K. W.; McKinnon, M.; Zhang, H.; Marathe, P. H.; Doweyko, A. M.; Kish, K.; Kiefer, S. E.; Sack, J. S.; Newitt, J. A.; Barrish, J. C.; Dodd, J.; Leftheris, K. The discovery of (*R*)-2-(sec-butylamino)-*N*-(2-methyl-5-(methylcarbamoyl)phenyl) thiazole-5-carboxamide (BMS-640994): A potent and efficacious p38α MAP kinase inhibitor. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1762−1767.

(32) Wrobleski, S. T.; Lin, S.; Dhar, T. G.; Dyckman, A. J.; Li, T.; Pitt, S.; Zhang, R.; Fan, Y.; Doweyko, A. M.; Tokarski, J. S.; Kish, K. F.; Kiefer, S. E.; Sack, J. S.; Newitt, J. A.; Witmer, M. R.; McKinnon, M.; Barrish, J. C.; Dodd, J. H.; Schieven, G. L.; Leftheris, K. The identification of novel p38α isoform selective kinase inhibitors having an unprecedented p38α binding mode. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 4120−4126.

(33) Weisberg, E.; Manley, P. W.; Breitenstein, W.; Bruggen, J.; Cowan-Jacob, S. W.; Ray, A.; Huntly, B.; Fabbro, D.; Fendrich, G.; Hall-Meyers, E.; Kung, A. L.; Mestan, J.; Daley, G. Q.; Callahan, L.; Catley, L.; Cavazza, C.; Azam, M.; Neuberg, D.; Wright, R. D.; Gilliland, D. G.; Griffin, J. D. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* **2005**, *7*, 129−141.

(34) Ishikawa, M.; Hashimoto, Y. Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry. *J. Med. Chem.* **2011**, *54*, 1539−1554.

(35) Pratt, E. F.; Kereszte, J. Syntheses of indolizino- and dihydroindolizinoquinoxalines. *J. Org. Chem.* **1967**, *32*, 49−53.

(36) Obafemi, C. A.; Pfleiderer, W. Synthesis and some reactions of 3-chloro-2-(cyanomethylene)-1,2-dihydroquinoxalines. *Molecules* **2004**, *9*, 223−231.

(37) Otomasu, H.; Ohmiya, S.; Sekuguch, T.; Takahash, H. Synthesis of condensed quinoxalines 0.2. A new synthesis of pyrrolo- 2,3-b-quinoxalines. *Chem. Pharm. Bull.* **1970**, *18*, 2065.

(38) Choi, H. G.; Ren, P.; Adrian, F.; Sun, F.; Lee, H. S.; Wang, X.; Ding, Q.; Zhang, G.; Xie, Y.; Zhang, J.; Liu, Y.; Tuntland, T.; Warmuth, M.; Manley, P. W.; Mestan, J.; Gray, N. S.; Sim, T. A type-II kinase inhibitor capable of inhibiting the T315I "gatekeeper" mutant of Bcr-Abl. *J. Med. Chem.* **2010**, *53*, 5439−5448.

(39) Huang, W. S.; Shakespeare, W. C. An efficient synthesis of nilotinib (AMN107). *Synthesis-Stuttgart* **2007**, 2121−2124.

(40) Jung, M. E.; Ouk, S.; Yoo, D.; Sawyers, C. L.; Chen, C.; Tran, C.; Wongvipat, J. Structure−activity relationship for thiohydantoin androgen receptor antagonists for castration-resistant prostate cancer (CRPC). *J. Med. Chem.* **2010**, *53*, 2779−2796.

(41) Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2007**, *2*, 2212−2221.

(42) Wang, Y.; Shakespeare, W. C.; Huang, W. S.; Sundaramoorthi, R.; Lentini, S.; Das, S.; Liu, S.; Banda, G.; Wen, D.; Zhu, X.; Xu, Q.; Keats, J.; Wang, F.; Wardwell, S.; Ning, Y.; Snodgrass, J. T.; Broudy, M. I.; Russian, K.; Dalgarno, D.; Clackson, T.; Sawyer, T. K. Novel N9-arenethenyl purines as potent dual Src/Abl tyrosine kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4907−4912.

(43) Bamborough, P.; Angell, R. M.; Bhamra, I.; Brown, D.; Bull, J.; Christopher, J. A.; Cooper, A. W.; Fazal, L. H.; Giordano, I.; Hind, L.; Patel, V. K.; Ranshaw, L. E.; Sims, M. J.; Skone, P. A.; Smith, K. J.; Vickerstaff, E.; Washington, M. *N*-4-Pyrimidinyl-1*H*-indazol-4-amine inhibitors of Lck: indazoles as phenol isosteres with improved pharmacokinetics. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4363−4368.

(44) Angell, R. M.; Angell, T. D.; Bamborough, P.; Brown, D.; Brown, M.; Buckton, J. B.; Cockerill, S. G.; Edwards, C. D.; Jones, K. L.; Longstaff, T.; Smee, P. A.; Smith, K. J.; Somers, D. O.; Walker, A. L.; Willson, M. Biphenyl amide p38 kinase inhibitors 2: Optimisation and SAR. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 324−328.

(45) Graham Robinett, R.; Freemerman, A. J.; Skinner, M. A.; Shewchuk, L.; Lackey, K. The discovery of substituted 4-(3-hydroxyanilino)-quinolines as potent RET kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5886−5893.

(46) Hynes, J., Jr.; Dyckman, A. J.; Lin, S.; Wrobleski, S. T.; Wu, H.; Gillooly, K. M.; Kanner, S. B.; Lonial, H.; Loo, D.; McIntyre, K. W.; Pitt, S.; Shen, D. R.; Shuster, D. J.; Yang, X.; Zhang, R.; Behnia, K.; Zhang, H.; Marathe, P. H.; Doweyko, A. M.; Tokarski, J. S.; Sack, J. S.; Pokross, M.; Kiefer, S. E.; Newitt, J. A.; Barrish, J. C.; Dodd, J.; Schieven, G. L.; Leftheris, K. Design, synthesis, and anti-inflammatory properties of orally active 4-(phenylamino)-pyrrolo[2,1-f][1,2,4]triazine p38α mitogen-activated protein kinase inhibitors. *J. Med. Chem.* **2008**, *51*, 4−16.

(47) Deak, H. L.; Newcomb, J. R.; Nunes, J. J.; Boucher, C.; Cheng, A. C.; DiMauro, E. F.; Epstein, L. F.; Gallant, P.; Hodous, B. L.; Huang, X.; Lee, J. H.; Patel, V. F.; Schneider, S.; Turci, S. M.; Zhu, X. *N*-(3-(Phenylcarbamoyl)arylpyrimidine)-5-carboxamides as potent and selective inhibitors of Lck: structure, synthesis and SAR. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1172−1176.

(48) Kufareva, I.; Abagyan, R. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. *J. Med. Chem.* **2008**, *51*, 7921−7932.

(49) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **2002**, *9*, 268−272.

(50) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discovery* **2006**, *5*, 730−739.

(51) Manley, P. W.; Cowan-Jacob, S. W.; Fendrich, G.; Jahnke, W.; Fabbro, D. Nilotinib, in Comparison to Both Dasatinib and Imatinib, Possesses a Greatly Prolonged Residence Time When Bound to the BCR-ABL Kinase SH1 Domain. 53rd ASH Annual Meeting and Exposition, San Diego, CA, 2011.

(52) Kitagawa, D.; Gouda, M.; Kirii, Y. Quick evaluation of kinase inhibitors by surface plasmon resonance using single-site specifically biotinylated kinases. *J. Biomol. Screening* **2014**, *19*, 453−461.

(53) Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **2004**, *64*, 6652−6659.

(54) Fabian, M. A.; Biggs, W. H., III; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.;

Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329−336.

(55) Kumar, S. R.; Singh, J.; Xia, G.; Krasnoperov, V.; Hassanieh, L.; Ley, E. J.; Scehnet, J.; Kumar, N. G.; Hawes, D.; Press, M. F.; Weaver, F. A.; Gill, P. S. Receptor tyrosine kinase EphB4 is a survival factor in breast cancer. *Am. J. Pathol.* **2006**, *169*, 279−293.

(56) Stephenson, S. A.; Slomka, S.; Douglas, E. L.; Hewett, P. J.; Hardingham, J. E. Receptor protein tyrosine kinase EphB4 is up-regulated in colon cancer. *BMC Mol. Biol.* **2001**, *2*, 15.

(57) Castellano, G.; Reid, J. F.; Alberti, P.; Carcangiu, M. L.; Tomassetti, A.; Canevari, S. New potential ligand-receptor signaling loops in ovarian cancer identified in multiple gene expression studies. *Cancer Res.* **2006**, *66*, 10709−10719.

(58) Adams, R. H. Vascular patterning by Eph receptor tyrosine kinases and ephrins. *Semin. Cell Dev. Biol.* **2002**, *13*, 55−60.

(59) Martiny-Baron, G.; Holzer, P.; Billy, E.; Schnell, C.; Brueggen, J.; Ferretti, M.; Schmiedeberg, N.; Wood, J. M.; Furet, P.; Imbach, P. The small molecule specific EphB4 kinase inhibitor NVP-BHG712 inhibits VEGF driven angiogenesis. *Angiogenesis* **2010**, *13*, 259−267.

(60) Korff, T.; Augustin, H. G. Tensional forces in fibrillar extracellular matrices control directional capillary sprouting. *J. Cell Sci.* **1999**, *112* (Pt 19), 3249−3258.

APPENDIX A: FRAGMENT-BASED DOCKING: DEVELOPMENT OF THE CHARMMING WEB USER INTERFACE AS A PLATFORM FOR COMPUTER-AIDED DRUG DESIGN

Pevzner, Y., Frugier, E., Schalk, V., Caflisch, A. and Woodcock, H.L., *J. Chem. Inf. Model.,* **2014**, 54(9), 2612-2620

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

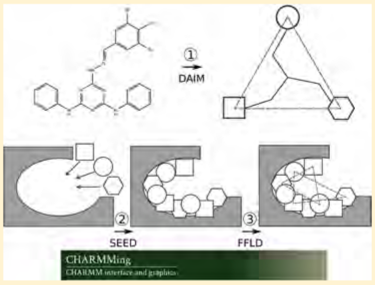
Article

pubs.acs.org/jcim

Fragment-Based Docking: Development of the CHARMMing Web User Interface as a Platform for Computer-Aided Drug Design

Yuri Pevzner,[†] Emilie Frugier,[‡] Vinushka Schalk,[†,§] Amedeo Caflisch,[‡] and H. Lee Woodcock*,[†]

[†]Department of Chemistry, University of South Florida, 4202 E. Fowler Ave., CHE205, Tampa, Florida 33620-5250, United States
[‡]Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland
[§]Department of Natural Sciences, New College of Florida, Sarasota, Florida 34243, United States



ABSTRACT: Web-based user interfaces to scientific applications are important tools that allow researchers to utilize a broad range of software packages with just an Internet connection and a browser.[1] One such interface, CHARMMing (CHARMM interface and graphics), facilitates access to the powerful and widely used molecular software package CHARMM. CHARMMing incorporates tasks such as molecular structure analysis, dynamics, multiscale modeling, and other techniques commonly used by computational life scientists. We have extended CHARMMing's capabilities to include a fragment-based docking protocol that allows users to perform molecular docking and virtual screening calculations either directly via the CHARMMing Web server or on computing resources using the self-contained job scripts generated via the Web interface. The docking protocol was evaluated by performing a series of "re-dockings" with direct comparison to top commercial docking software. Results of this evaluation showed that CHARMMing's docking implementation is comparable to many widely used software packages and validates the use of the new CHARMM generalized force field for docking and virtual screening.

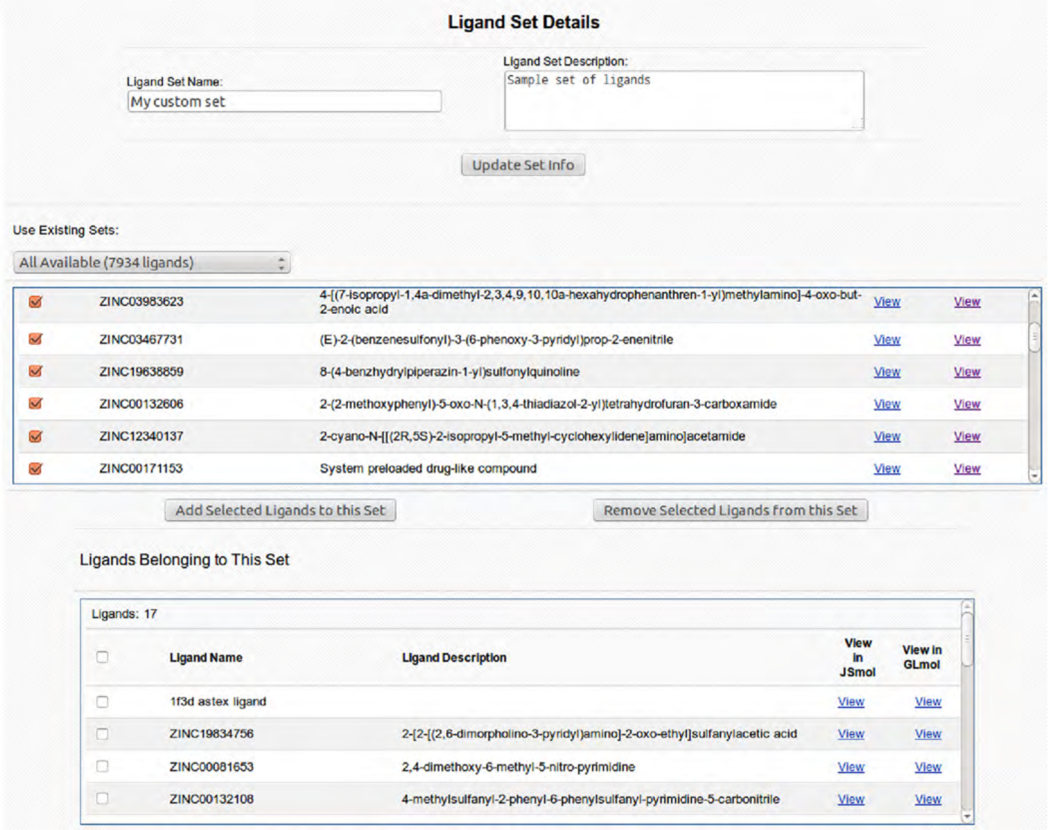
■ INTRODUCTION

In the past, the cost and effort of developing a new drug has largely confined successes to large pharmaceutical companies or otherwise well-funded research institutions.[2] Although development and use of computer-aided drug design (CADD) techniques has provided numerous benefits to the overall process, the expertise required to create powerful commercial software packages has resulted in high licensing costs,[3,4] thus limiting access to academic groups. Fortunately, this trend has started to shift with the emergence of freely available software, such as Autodock[5] and several other packages,[4] largely developed by the academic computational chemistry community. However, for the most part, these software packages require familiarity with CADD methodologies and are better suited for computer savvy users that are at least comfortable if not familiar with the computational component of drug discovery.[6] This has hampered the proliferation of CADD tools into less computationally minded drug discovery laboratories. The need for intuitive and easy to use CADD solutions has largely been met by the commercial software companies such as Accelrys, Schrödinger, and others that have incorporated full-featured graphical user interfaces (GUI) into their programs.[7−9] However, as alluded to above, the cost of these packages is typically prohibitive to academic groups and/or institutions. Further, it has proven increasingly difficult to strike a balance between software that is user-friendly yet incorporates a wide range of advanced functionality and customizability. Another aspect of concern is portability. For example stand-alone software that requires local installation on every computer may find less use in today's world where researchers expect both the application and the data to be accessible from any machine on any platform from any location.[10]

Another hurdle, faced by the nonexpert, to incorporating computational modeling into drug discovery efforts is the difficulty of obtaining reliable small molecule parameters.[11−13] Most widely used and well-tested force fields have been developed with proteins and nucleic acids rather than small molecules in mind.[14] Until recently this has meant that drug-like molecule parameters have been less reliable, with assignment often arbitrary. Lately, however, there has been a significant amount of effort devoted to improving the reliability of small molecule parameters and developing efficient protocols to generate them for a much greater and more diverse chemical space.[11,12,14,15]

The CHARMM interface and graphics (CHARMMing)[16] is a Web interface to the popular macromolecular modeling package CHARMM.[17,18] The goal of the CHARMMing project is to provide a platform-independent Web-based front-end that allows its users to set up and perform a wide variety of molecular modeling tasks. CHARMMing's users range from small academic laboratories that benefit from the portal's functionality to educators that include molecular modeling in their curricula and use the portal to facilitate their teaching.[19−21] Moreover, the open source nature of the project allows outside developers to utilize the framework and build on

Received: May 30, 2014
Published: August 24, 2014



Ligand Set Details

Ligand Set Name: My custom set

Ligand Set Description: Sample set of ligands

Use Existing Sets: All Available (7934 ligands)

☑	ZINC03983623	4-[(7-isopropyl-1,4a-dimethyl-2,3,4,9,10,10a-hexahydrophenanthren-1-yl)methylamino]-4-oxo-but-2-enoic acid	View	View
☑	ZINC03467731	(E)-2-(benzenesulfonyl)-3-(6-phenoxy-3-pyridyl)prop-2-enenitrile	View	View
☑	ZINC19638859	8-(4-benzhydrylpiperazin-1-yl)sulfonylquinoline	View	View
☑	ZINC00132606	2-(2-methoxyphenyl)-5-oxo-N-(1,3,4-thiadiazol-2-yl)tetrahydrofuran-3-carboxamide	View	View
☑	ZINC12340137	2-cyano-N-[[(2R,5S)-2-isopropyl-5-methyl-cyclohexylidene]amino]acetamide	View	View
☑	ZINC00171153	System preloaded drug-like compound	View	View

Ligands Belonging to This Set

Ligands: 17

☐	Ligand Name	Ligand Description	View in JSmol	View in GLmol
☐	1f3d astex ligand		View	View
☐	ZINC19834756	2-[2-[(2,6-dimorpholino-3-pyridyl)amino]-2-oxo-ethyl]sulfanylacetic acid	View	View
☐	ZINC00081653	2,4-dimethoxy-6-methyl-5-nitro-pyrimidine	View	View
☐	ZINC00132108	4-methylsulfanyl-2-phenyl-6-phenylsulfanyl-pyrimidine-5-carbonitrile	View	View

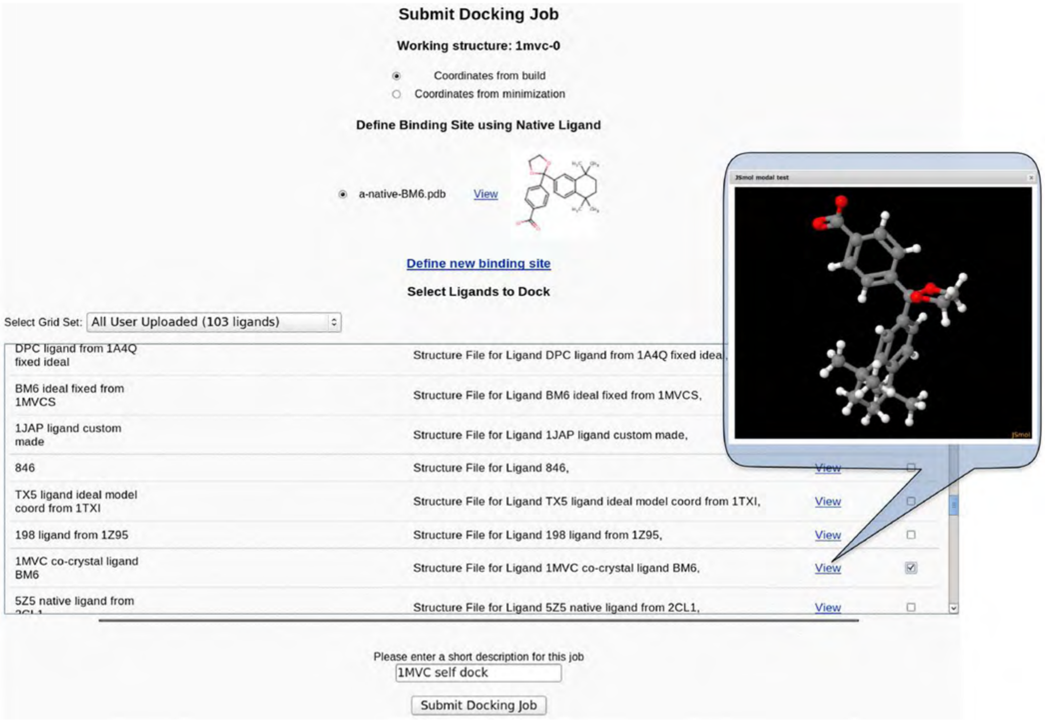
Figure 1. "Ligand Set Details" page allows the user to manage custom ligand sets. The user can define and describe a custom ligand set as well as add ligands to it from any of the other sets including the preloaded public library.

the existing infrastructure, further expanding the range of features it includes. The framework can be easily installed on a private network or adopted into a new Web-based interface; this approach was utilized when developing a virtual target screening (VTS)[22] server. Herein, we describe a similar effort using the CHARMMing infrastructure (i.e., built on a Python-based[23] Django[24] framework with a MySQL[25] database); the implementation of a new drug design module that incorporates a fragment-based docking protocol includes a diverse set of drug-like compounds and facilitates creation of CHARMM friendly protein–small molecule systems for further modeling studies. We also assess the performance of the newly implemented docking protocol coupled to CHARMM's new generalized force field (CGenFF) by reproducing a series of co-crystallized protein–ligand complexes and comparing the results against a leading commercially available docking package.

■ IMPLEMENTATION DETAILS

Target Preparation. Target proteins begin their preparation via CHARMMing's structure submission section. Here, tasks such as the addition of hydrogens, identification of any nonprotein moieties, and assignment of final parameters are carried out (using the latest CHARMM36 protein force field).[26,27] Co-crystallized small molecules (i.e., ligands) are automatically parametrized using the CGenFF.[12] Specifically, ligand atom-typing and parametrization is performed by sequentially attempting several automated parametrization tools. The default order is (1) ParamChem,[12,28,29] (2) MATCH,[30] (3) Antechamber,[31] and (4) GENRTF.[32] As an alternative to the default order, a user can specify the exact build procedure to use for parametrization.

Compound Library and Ligand Upload. CHARMMing docking module provides a preloaded library of drug-like compounds for virtual screening experiments. The library consists of approximately 8000 molecules from the Maybridge Hitfinder set (www.maybridge.com). All of the provided molecules have been atom typed according to CGenFF convention to comply with CHARMM requirements and confirmed to decompose into at least three sufficiently sized fragments to meet the fragment-based docking criteria. CHARMMing also allows users to upload ligands by providing a coordinate file in mol2 format. Upon uploading, the ligand undergoes atom-typing and parametrization as previously described. The ligand and corresponding parameter, topology, and structure files are then saved on a disk as well as cataloged



Submit Docking Job

Working structure: 1mvc-0

● Coordinates from build
○ Coordinates from minimization

Define Binding Site using Native Ligand

● a-native-BM6.pdb View

Define new binding site

Select Ligands to Dock

Select Grid Set: All User Uploaded (103 ligands)

DPC ligand from 1A4Q fixed ideal	Structure File for Ligand DPC ligand from 1A4Q fixed ideal,		
BM6 ideal fixed from 1MVCS	Structure File for Ligand BM6 ideal fixed from 1MVCS,		
1JAP ligand custom made	Structure File for Ligand 1JAP ligand custom made,		
846	Structure File for Ligand 846,	View	☐
TX5 ligand ideal model coord from 1TXI	Structure File for Ligand TX5 ligand ideal model coord from 1TXI,	View	☐
198 ligand from 1Z95	Structure File for Ligand 198 ligand from 1Z95,	View	☐
1MVC co-crystal ligand BM6	Structure File for Ligand 1MVC co-crystal ligand BM6,	View	☑
5Z5 native ligand from	Structure File for Ligand 5Z5 native ligand from 2CL1,	View	☐

Please enter a short description for this job
1MVC self dock

Submit Docking Job

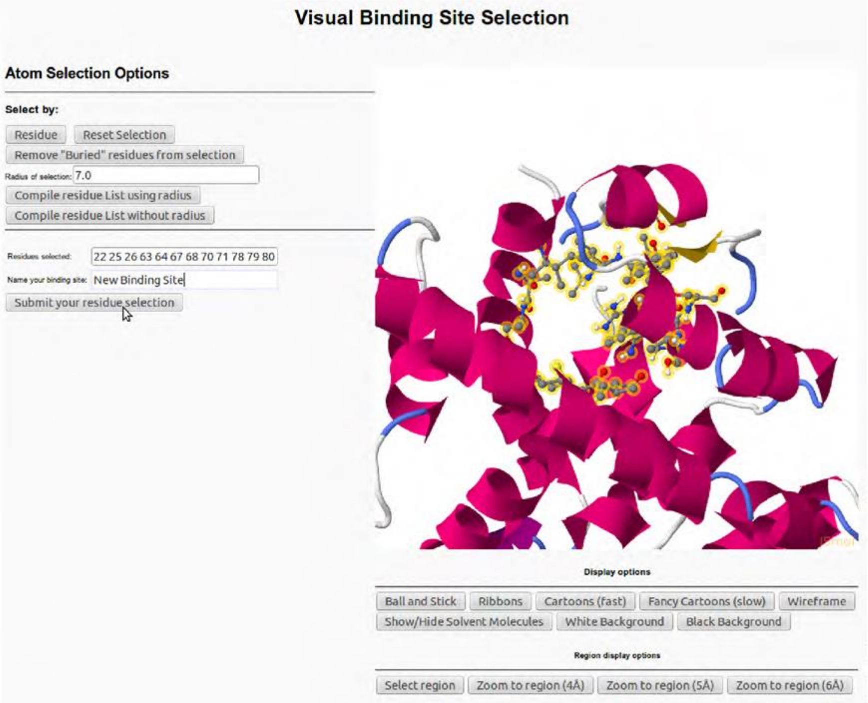
Figure 2. "Submit Docking Job" page presents the user with the ability to select the target coordinates for docking, define the binding pocket (*vide infra*), and select ligands to dock from the list of available small molecules. Native ligands and ligands available for docking can be visualized in 3D using the embedded visualization application.

in the database. Unlike the preloaded compound library, any user-uploaded ligands are restricted to their account only and are not visible to other users. The user is also given the ability to create custom sets of molecules based on any preloaded or user-uploaded compounds. This can be done via the "Ligand Sets" section (Figure 1) of the docking module. Any custom or preloaded set can be docked in its entirety or by selecting individual molecules on the docking submission page (Figure 2).

Binding Site Definition. To provide maximum flexibility with respect to job setup, two different ways of specifying the binding region of interest are implemented. The first approach identifies the binding pocket using the position of a co-crystallized ligand that may be present. In this case, when launching a docking job, a user is presented with a list of all co-crystallized small molecules along with their 2D structural representations. Once the desired small molecule is chosen, the binding site is defined via proximity to the aforementioned small molecule. In cases where no co-crystallized ligand is present, or if a user simply wishes to investigate alternative binding sites, we have implemented an interactive and graphical binding site definition tool (Figure 3). To use this tool, two residues should be selected that roughly correspond to the edges of the desired binding region. The midpoint between these residues is then determined and defined as the approximate center of the binding site. On the basis of a user-defined radius, a list of all residues within this distance is compiled and both visually highlighted and presented as a list. The user can then add or remove residues to/from this list by either modifying the text of the residue list, changing the specified search radius, or modifying it via graphical selection (i.e., clicking). Ultimately, all user-defined binding sites are saved and presented as options with any existing co-crystallized ligands at the docking job setup page.

Docking Protocol. Docking algorithms used in this protocol are based on the popular grid-based paradigm used by most current docking programs.[33−38] In this approach, the solvent accessible surface area of the target and the ligand as well as the target's binding site are discretized onto a 3D lattice. The lattice then either stores information about the atoms enclosed by a cubic unit of the grid or contains the potential contributions projected onto the grid's vertices. Precomputed grids allow for efficient calculation of both van der Waals and electrostatic contributions to the scoring function, facilitating rapid evaluation of ligand placements within the binding site.

The docking procedure consists of several steps where different programs perform distinct tasks. To streamline the communication between the programs and ensure compatibility of input and output data, a series of scripts were written in Python, Perl, and Linux shell scripting languages. The OpenBabel[39] file conversion utility was used to interconvert between different representations of the protein and compound structures. The program MATCH[30] was used to generate CGenFF compatible topologies and parameters. The fragment-



Visual Binding Site Selection

Atom Selection Options

Select by:

Radius of selection: 7.0

Residues selected: 22 25 26 63 64 67 68 70 71 78 79 80

Name your binding site: New Binding Site

Figure 3. "Binding Site Selection" page provides the user multiple ways to select a custom binding site. This can be done either by manually typing in the residue numbers, graphically selecting residues, or defining the centroid and specifying the radius in Å.

based docking protocol implemented in CHARMMing is outlined in Figure 4 and described as follows:

(1) Each compound to be docked is first broken down into fragments. A fingerprint describing chemical richness is generated for each fragment and its parent compound. The three most chemically rich, but not necessarily different, fragments are identified to serve as anchors for docking. These steps are carried out by the program DAIM (Decomposition and Identification of Molecules).[40]

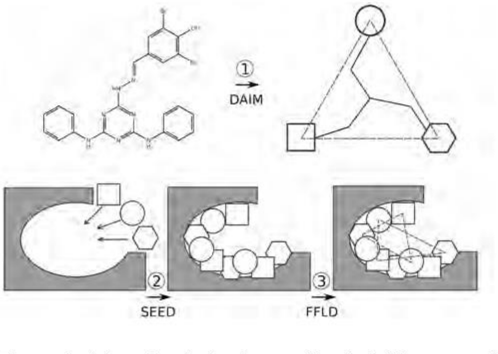


Figure 4. Schematic of the fragment-based docking protocol implemented into the CHARMMing Web user interface. Depicted are the three main stages of the docking: decomposition by DAIM, fragment docking by SEED, and ligand placement by FFLD.

(2) The user then identifies the binding site to be used in the docking job. All nonprotein nonsolvent compounds present in the submitted target structure are displayed on the "Submit Docking Job" page (Figure 2). On the basis of the user selected compound, the proximal residues are identified and the binding site defined.

(3) The previously identified anchor fragments (step 1) are then docked into the binding site using the program SEED (Solvation Energy for Exhaustive Docking).[41] The placement of fragments within the binding site is determined by matching either the direction of polar vectors between ligand and receptor atoms to form a hydrogen bond or the apolar vectors on the solvent accessible surface area of the receptor. The SEED score, used in fragment placement, accounts for the solvent effects by including terms for both receptor and fragment desolvation as well as a solvent-screened receptor-fragment electrostatic interaction term.

(4) The docked fragments are reconnected into the original ligand while undergoing refinement using the FFLD (Fragment-based Flexible Ligand Docking) program.[42] FFLD uses a genetic algorithm that generates and evaluates populations of conformations and positions them within the binding site, as guided by fragment anchor locations. The fitness of a placed conformation is evaluated using a scoring function that is aimed at approximating the steric effects as well as hydrogen bonding contributions of the protein−ligand interactions. This function

Job Details

Job ID	Description	Current Status	Submitted	Wait time	Running time	Total time	Results
106	1MVC self dock	Done	May 15, 2014, 10:44 a.m.	None	None	0:04:33	View/Hide Results

Results

Rank	Ligand Name	DSFC Average	DSFC Median	CHARMM Energy	SEED Score	FFLD Score	Download File	View File	View Pose
1	1MVC co-crystal ligand BM6	-60.4117333333	-48.88	-150.294	-55.5364	-42.51	Download File	View File	View Pose

Ligand Name	Ligand Description	View File	Download File
1MVC co-crystal ligand BM6		View File	Download File

File Description	View File	Download File
Archived Downloadable Drug Design Job 106	View File	Download File
Job execution output for job dd_job_106	View File	Download File

Figure 5. "Job Details" page provides general job information as well as the list of docked poses and their respective scores. The docked poses can be visualized in 3D within the binding pocket of the protein using the embedded visualization application. An archive of the job directory can also be downloaded from this page for execution on local resources.

includes intraligand and protein–ligand van der Waals interaction terms as well as polar contributions based on the number of hydrogen bonds and unfavorable donor–donor and acceptor–acceptor interactions.

(5) Poses generated by FFLD that are within a user-defined energy cutoff (10 kcal/mol by default) are then clustered using a leader clustering algorithm implemented in the program FLEA (FFLD Leader Clustering).[43]

(6) Following the clustering, the protein–ligand complex is converted to native CHARMM format and saved. Using these files, in addition to the CHARMM protein and generalized force fields (i.e., CHARMM36 and CGenFF), protein structure (psf) and coordinate (crd) files are generated. Each ligand then undergoes 1000 steps of minimization using the adopted Newton–Rhapson (ABNR) algorithm while keeping protein atoms fixed. The "minimized" protein–ligand complexes are then scored using SEED and FFLD in their "evaluation only" mode, producing their own estimation of electrostatic, van der Waals, and total energy contributions for each pose. The final ranking of the docked poses is performed using a consensus approach. For this, energies (i.e., interaction energy from CHARMM and total energies from SEED and FFLD) are used to create three lists in which individual poses are sorted and ranked. The final rank of each pose is then set to the median of the three ranks as assigned in the individual lists. The consensus approach to scoring or ranking compounds when performing molecular docking or virtual screening studies has been shown to be more accurate than single scoring methods.[44−49]

Job Submission and Monitoring. When a docking job is launched, the PBS[50] (Portable Batch System)-based queuing system TORQUE[51] accepts the job as a wrapper shell script that controls the entire docking procedure. Using the interface, a job can be monitored in real time as it progresses and generates final poses for each docked compound. Basic job statistics such as submission time and job status can be monitored along with the output file reflecting the job progression (Figure 5). In addition, important files associated with job progress and results (e.g., final docked ligand poses, job output, etc.) can be downloaded to a local disk. Protein, ligands, compounds in the library, and final docked poses can all be visualized directly in CHARMMing. The 3D structure of each of the above elements can be rendered with the JSmol[52] or GLmol[53] visualization tools. Structures can be visualized using a variety of representations to highlight important structural features or interactions of the molecules and their complexes.

A walk-through outlining the entire process of performing a redock on a sample system is included in the tutorial covering basic CHARMM and CHARMMing functionality at www.charmmtutorial.org. Additionally, a docking lesson that guides a user through the redocking procedure has been added to the lessons section of the CHARMMing Web site.

Performance and Local Execution. Currently, all docking jobs executed via the Web interface are carried out sequentially. However, after initial setup of the docking job, all necessary files are available for download and execution on local computational resources. To improve performance of this procedure, we have developed a protocol that can be carried out in parallel as outlined in Figure 6. This is achieved by spawning a new execution branch for each of the most time-consuming steps in the protocol via a user-modifiable job queuing command. For example, each fragment of each molecule is docked (step 3, vide supra) as a separate submitted job. Once all of a molecule's anchor fragments are docked, the placement of a ligand within the binding site by FFLD is also spawned as a series of separate jobs. Furthermore, to increase sampling by FFLD and improve performance, the protocol performs multiple docking iterations per ligand, again each as a separate job. Thus, instead of one docking job that attempts to sequentially sample a large conformational space per ligand, multiple shorter iterations with different random seeds are run in parallel, taking less real time and still sufficiently sampling ligand conformational space. The number of iterations per ligand as well as the amount of energy evaluations per iteration are all user modifiable parameters.

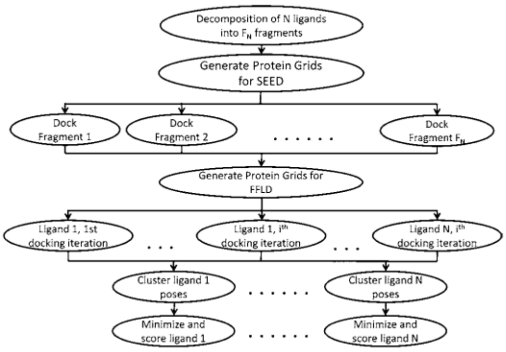


Figure 6. Parallelization of the docking protocol is achieved by spawning new job execution threads at both the fragment docking (i.e., one per fragment) and ligand placement (i.e., one per iteration per ligand) steps. Clustering and scoring threads are also spawned for each docked ligand.

In order to execute a job on local resources, the following programs need to be downloaded and installed: VMD,[54] DAIM, SEED, FFLD, FLEA, MATCH, and CHARMM. Except for CHARMM, all of these programs are free for academic use. VMD can be downloaded from the University of Illinois at Urbana−Champaign's Theoretical and Computational Biophysics group (www.ks.uiuc.edu/Research/vmd). DAIM, SEED, FFLD, and FLEA can be obtained from the University of Zurich's Computational Structural Biology lab (www.biochem-caflisch.uzh.ch/download). Further, a more general description of the installation process is included as part of the CHARMM tutorial and can be found at the following Web address: www.charmmtutorial.org/index.php/Installation_of_CHARMMing.

Once the job directory is downloaded and the software is installed on local resources, the provided settings file should be used to specify the location of program executables. In addition, job details (e.g., protein file name, number of docking iterations, clustering energy cutoff, etc.) can be modified via the settings file. This file is also where PBS/TORQUE commands can be modified for local resources. Because there is no limit to the number of possible parallel processes spawned, the protocol checks for available resources and will wait for current processes to complete if the queue is full. The protocol will automatically take advantage of all available resources to speed up job completion while at the same time adhering to the local queuing system policies.

■ RESULTS AND DISCUSSION

To assess the performance of the docking protocol, a diversity set was constructed from the publicly available CCDC/Astex test set[55] containing high-resolution X-ray complexes and an augmented version of that set, which has been used to compare the performance of a number of docking programs.[56] Our final set contained 24 protein−ligand complexes with X-ray resolutions ranging from 1.50−2.30 Å. In particular, we selected complexes where the ligand could be decomposed into three fragments (i.e., at least three rotatable bonds) using the default settings of DAIM, as the ultimate goal was to evaluate the implementation of the decomposition-based approach.

Redock validation involved removing the co-crystallized ligand from the complex, redocking it via the fragment-based protocol, and comparing the docked pose to that of the original crystal structure. Each complex was processed using CHARMMing's "Submit Structure" section that downloads the structure based on the PDB code, adds hydrogen atoms, and prepares the structure for modeling using CHARMM. Further, each system containing the protein, solvent, and ligand molecules was briefly minimized for 100 steps using the Steepest Descent method followed by 1000 steps of ABNR using CHARMMing's "Calculations" module. Using the "Ligand Upload" section of CHARMMing's docking module, the previously downloaded ligand was processed. The docking calculation for each minimized system was set up by selecting a native ligand to define a binding pocket and user-uploaded ligand for docking, all from the "Submit Docking Job" page of the docking module. The progress of each job was monitored using the job monitoring section of the docking module. To assess the performance of the dockings, root-mean-square deviation (RMSD) between the heavy atoms of the docked poses and the crystal structures was calculated using VMD.

To compare the docking protocol's performance, a commercially available docking package was also used. Redockings were performed using Schrödinger's Glide[34−36,57] Standard Precision (SP) docking protocol. Glide's SP protocol attempts to dock multiple conformations of a ligand into a receptor grid, subsequently calculating the effective ligand−receptor interactions using a proprietary scoring function. Conformational sampling of the ligand is achieved via varying torsion angles around rotatable bonds. Prior to docking, each target was prepared using Maestro's[7] Protein Preparation Wizard.[58−62] The preparation included removal of solvent molecules, addition of hydrogens, and brief minimization. As Glide is also a grid-based docking protocol, the grids, similar to CHARMMing's procedure, were built using the co-crystal ligand to define the binding region. The native ligand was removed and redocked using default parameters of the SP docking protocol. The poses with the best docking scores were used to calculate their respective RMSD from the crystal structure using VMD.

Table 1 reports the RMSD of poses generated by CHARMMing's fragment-based docking protocol and Glide SP docking (w.r.t. crystal structure). Results reported from CHARMMing's fragment-based docking protocol correspond to the pose closest to the crystal structure. This set yields a 71% success rate using RMSD < 2.0 Å as the metric; this criteria is commonly employed for evaluating the performance of docking algorithms.[56,63−65] This clearly shows that the protocol can successfully recover the crystal pose in the majority of the cases. We are currently optimizing a consensus scoring function based on this diversity set; results of that effort will be reported in a subsequent publication. Nevertheless, virtual screening is known to suffer from a high false-positive rate, which does not diminish its value in drug discovery as the unfit compounds are screened out during the experimental stages of the discovery campaigns.[66] Regardless, we are encouraged by the success of fragment-based docking, which shows approximately the same performance as widely used docking programs, i.e., within the range of 40−90%.[56,63−65]

Table 1. RMSDs of Docking Poses Generated by CHARMMing's Fragment-Based Docking Protocol and Glide SP and success Rates (defined by the percentage of the ligands whose reported RMSD is below 2.0 Å)[a]

PDB ID	resolution (Å)	best RMSD (Å)	Glide SP RMSD (Å)
1A4Q	1.90	2.61	3.30
1A6W	2.00	1.01	6.72
1AOE	1.60	3.13	1.80
1AQW	1.80	1.88	0.96
1ATL	1.80	1.68	1.09
1BMA	1.80	2.76	1.55
1D3H	1.80	0.99	0.81
1FCZ	1.38	1.06	0.31
1GLQ	1.80	4.71	1.01
1HFC	1.50	2.63	2.36
1HVR	1.80	3.84	0.75
1JAP	1.82	1.41	0.92
1KES	2.20	1.07	1.75
1MLD	1.83	1.29	1.07
1MMQ	1.90	0.50	0.30
1MTS	1.90	1.96	0.54
1MVC	1.90	0.29	0.94
1NHZ	2.30	0.78	1.89
1NQ7	1.50	0.94	1.26
1QBR	1.80	9.31	0.98
1SRJ	1.80	1.33	0.51
1TXI	1.90	1.66	1.64
3ERT	1.90	0.71	1.59
4DFR	1.70	1.66	10.48
Success Rate:		71%	84%

[a] "Best RMSD" refers to the pose closest to the crystal structure. Glide SP RMSD is of the top scoring pose of Glide's standard precision docking.

The fragment-based approach that was implemented into CHARMMing yields a substantial amount of information about the characteristics of each docked pose. At each step, from decomposition to minimization of docked poses, users have the ability to closely analyze results. The binding modes of each individual fragment can be inspected, and a number of modifiable parameters, such as decomposition criteria, can be used to optimize the protocol. Moreover, information gained from docking a fragment library into a particular target can be used to mine large libraries for compounds containing those fragments that form the most favorable interactions with the target.[67−69]

There are potentially a number of improvements that can be made to improve the performance and usability of CHARMMing's docking protocol. The most obvious limitation is the current requirement of three fragments to be used as anchors. As shown by the number of ligands eliminated from the original benchmarking set, this limits the applicability of this protocol in its current form to medium- to large-sized molecules with a sufficient number of rotatable bonds. Although partially this problem can be alleviated by decreasing the fragment richness threshold at the decomposition step, this will only increase the "eligibility" rate of molecules by a small margin. Alternatively, when docking these small and/or rigid molecules is desired, the decomposition step could be omitted, at which point the molecules would undergo docking only by SEED. This however will require prior conformation sampling step as SEED currently does not sample the internal conformation of docked fragments. The conformational sampling of the fragments is an obvious improvement to the docking protocol even in its current state. This addition will help ensure that larger fragments sample their orientations within the binding site while varying their internal geometry, thus ensuring greater enrichment of anchor positions for the final ligand placement. Efforts to incorporate these functionality improvements are currently underway.

■ CONCLUSIONS

We have implemented a fragment-based docking protocol into the CHARMMing Web interface. The protocol allows users to perform docking and virtual screening calculations online as well as generates self-contained scripts to execute these in parallel on local HPC resources. The performance of the docking protocol was evaluated by carrying out a series of redockings and comparing the results against a top commercial docking package. The fragment-based docking protocol yielded results comparable to both the commercial package used herein and a wide variety of additional docking software. Specifically, the rate of recovering the correct X-ray pose with CHARMMing's protocol was 71%, well within the 40−90% range that numerous benchmarking studies have reported.

While the scoring function can still be improved, the tool lays substantial groundwork for allowing academic laboratories to set up and perform molecular docking and virtual screening studies. It is important to note that the protocol is able to create CHARMM-formatted protein−ligand systems giving users the ability to access the wide range of functionality that exists in CHARMM. For example, docked poses can easily be refined with MD simulations, and predocked proteins can be coupled with simulations or normal-mode analysis to proceed via an ensemble docking approach. These, in addition to other improvements are currently being developed.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hlw@usf.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

H.L.W. would acknowledges the NIH (1K22HL088341-01A1), DOE (DE-SC0011297TDD), NSF (CHE1156853) and the University of South Florida (start-up) for funding. Computations were performed at the USF Research Computing Center. The authors thank Dr. Sandra Rennebaum for her assistance with troubleshooting docking programs used in the protocols, Dr. Peter Kolb for help with DAIM, Prof. Alex MacKerell and Dr. Kenno Vanommeslaeghe for assistance with small molecule atom-typing and access to the Maybridge library. Finally, we especially thank Mr. Tim Miller, Dr. Bernard Brooks, and the National Heart, Lung, and Blood Institute of the NIH for support and collaboration.

■ REFERENCES

(1) Bello, M.; Martínez-Archundia, M.; Correa-Basurto, J. Automated docking for novel drug discovery. *Expert Opin. Drug Discovery* **2013**, *8*, 821–834.

(2) Moors, E. H. M.; Cohen, A. F.; Schellekens, H. Towards a sustainable system of drug development. *Drug Discovery Today* **2014**, 1–10.

(3) DeLano, W. L. The case for open-source software in drug discovery. *Drug Discovery Today* **2005**, *10*, 213–217.

(4) Geldenhuys, W. J.; Gaasch, K. E.; Watson, M.; Allen, D. D.; Van der Schyf, C. J. Optimizing the use of open-source software applications in drug discovery. *Drug Discovery Today* **2006**, *11*, 127–132.

(5) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195–202.

(6) Lill, M. A.; Danielson, M. L. Computer-aided drug design platform using PyMOL. *J. Comput.-Aided. Mol. Des.* **2011**, *25*, 13–19.

(7) *Maestro*; Schrödinger, LLC: New York, 2013.

(8) *Discovery Studio Modeling Environment*; Accelrys Software, Inc.: San Diego, CA, 2013.

(9) *Molecular Operating Environment (MOE), 2013.08*; Chemical Computing Group, Inc.: Montreal, Quebec, Canada, 2013.

(10) Ebejer, J.-P.; Fulle, S.; Morris, G. M.; Finn, P. W. The emerging role of cloud computing in molecular modelling. *J. Mol. Graph. Model.* **2013**, *44*, 177–187.

(11) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

(12) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

(13) Knight, J. L.; Brooks, C. L. Validating CHARMM parameters and exploring charge distribution rules in structure-based drug design. *J. Chem. Theory Comput.* **2009**, *5*, 1680–1691.

(14) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(15) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **2011**, *32*, 2359–2368.

(16) Miller, B. T.; Singh, R. P.; Klauda, J. B.; Hodoscek, M.; Brooks, B. R.; Woodcock, H. L. CHARMMing: A new, flexible web portal for CHARMM. *J. Chem. Inf. Model.* **2008**, *48*, 1920–1929.

(17) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(18) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(19) Miller, B. T.; Singh, R. P.; Schalk, V.; Pevzner, Y.; Sun, J.; Miller, C. S.; Boresch, S.; Ichiye, T.; Brooks, B. R.; Woodcock, H. L., III Web-based computational chemistry education with CHARMMing I: Lessons and tutorial. *PLoS Comput. Biol.* **2014**, *10*, e1003719.

(20) Pickard, F. C., IV; Miller, B. T.; Schalk, V.; Lerner, M. G.; Woodcock, H. L., III; Brooks, B. R. Web-based computational chemistry education with CHARMMing II: Coarse-grained protein folding. *PLoS Comput. Biol.* **2014**, *10*, e1003738.

(21) Perrin, B. S., Jr; Miller, B. T.; Schalk, V.; Woodcock, H. L.; Brooks, B. R.; Ichiye, T. Web-based computational chemistry education with CHARMMing III: Reduction potentials of electron transfer proteins. *PLoS Comput. Biol.* **2014**, *10*, e1003739.

(22) Pevzner, Y.; Santiago, D. N.; Fries, J. S.; Metcalf, R. S.; Daniel, K. G.; Calcul, L.; Woodcock, H. L.; Baker, B. J.; G, W. C.; Brooks, W. H. Virtual target screening to rapidly identify potential protein targets of natural products in drug discovery. *AIMS Mol. Sci.* **2014**, *1*, 49–66.

(23) Python, Python Software Foundation. www.python.org (accessed August 27, 2014).

(24) Django, Django Software Foundation. http://www.djangoproject.com (accessed August 27, 2014).

(25) *MySQL*; Oracle Corporation: Redwood Shores, CA, 2014.

(26) MacKerell, A. D.; Bashford, D.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(27) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(28) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

(29) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

(30) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L. MATCH: An atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem.* **2012**, *33*, 189–202.

(31) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(32) Hodoscek, M. *GENRTF*, 2008. http://code.google.com/p/genrtf/ (accessed August 27, 2014).

(33) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.

(34) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(35) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(36) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(37) Grosdidier, A.; Zoete, V.; Michielin, O. Fast docking using the CHARMM force field with EADock DSS. *J. Comput. Chem.* **2011**, 2149−2159.

(38) Wu, G.; Robertson, D. H.; Brooks, C. L.; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549−1562.

(39) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33−47.

(40) Kolb, P.; Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.* **2006**, *49*, 7384−7392.

(41) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins* **2001**, *42*, 256−268.

(42) Budin, N.; Majeux, N.; Caflisch, A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.* **2001**, *382*, 1365−1372.

(43) Dey, F.; Caflisch, A. FLEA − FFLD Leader Clustering. http://www.biochem-caflisch.uzh.ch/download/ (accessed August 27, 2014).

(44) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.

(45) Ece, A.; Sevin, F. The discovery of potential cyclin A/CDK2 inhibitors: A combination of 3D QSAR pharmacophore modeling, virtual screening, and molecular docking studies. *Med. Chem. Res.* **2013**, *22*, 5832−5843.

(46) Ji, X.; Zheng, Y.; Wang, W.; Sheng, J.; Hao, J.; Sun, M. Virtual screening of novel reversible inhibitors for marine alkaline protease MP. *J. Mol. Graph. Model.* **2013**, *46*, 125−131.

(47) Houston, D. R.; Walkinshaw, M. D. Consensus docking: Improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model.* **2013**, *53*, 384−390.

(48) Friedman, R.; Caflisch, A. Discovery of plasmepsin inhibitors by fragment-based docking and consensus scoring. *ChemMedChem.* **2009**, *4*, 1317−1326.

(49) Klon, A. E.; Glick, M.; Davies, J. W. Combination of a naive Bayes classifier with consensus scoring improves enrichment of high-throughput docking results. *J. Med. Chem.* **2004**, *47*, 4356−4359.

(50) OpenPBS, PBS Works. www.openpbs.org (accessed August 27, 2014).

(51) TORQUE, Cluster Resources, Inc. (Adaptive Computing, Inc.).http://www.clusterresources.com (accessed August 27, 2014).

(52) Hanson, R. M.; Prilusky, J.; Renjian, Z.; Nakane, T.; Sussman, J. L. JSmol and the next-generation Web-based representation of 3D molecular structure as applied to proteopedia. *Isr. J. Chem.* **2013**, *53*, 207−216.

(53) Nakane, T. GLmol − Molecular Viewer on WebGL/Javascript, version 0.47. http://webglmol.sourceforge.jp/index-en.html (accessed August 27, 2014).

(54) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(55) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein−ligand interaction. *Proteins* **2002**, *49*, 457−471.

(56) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455−1474.

(57) *Small-Molecule Drug Discovery Suite 2014−1: Glide*; Schrödinger, LLC: New York, 2014.

(58) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided. Mol. Des.* **2013**, *27*, 221−234.

(59) *Epik*; Schrödinger, LLC: New York, 2013.

(60) *Impact*; Schrödinger, LLC: New York, 2013.

(61) *Prime*; Schrödinger, LLC: New York, 2014.

(62) *Protein Preparation Wizard*, Schrödinger Suite 2013; Schrödinger, LLC: New York, 2013.

(63) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *56*, 235−249.

(64) Srivastava, H. K.; Chourasia, M.; Kumar, D.; Sastry, G. N. Comparison of computational methods to model DNA minor groove binders. *J. Chem. Inf. Model.* **2011**, *51*, 558−571.

(65) Neale, D. S.; Thompson, P. E.; White, P. J.; Chalmers, D. K.; Yuriev, E.; Manallack, D. T. Binding mode prediction of PDE4 inhibitors: A comparison of modelling methods. *Aust. J. Chem.* **2010**, *63*, 396−404.

(66) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862−865.

(67) Kolb, P.; Kipouros, C. B.; Huang, D.; Caflisch, A. Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins* **2008**, *73*, 11−18.

(68) Zhao, H.; Dong, J.; Lafleur, K.; Nevado, C.; Caflisch, A. Discovery of a novel chemotype of tyrosine kinase inhibitors by fragment-based docking and molecular dynamics. *ACS Med. Chem. Lett.* **2012**, *3*, 834−838.

(69) Zhao, H.; Gartenmann, L.; Dong, J.; Spiliotopoulos, D.; Caflisch, A. Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2493−2496.

APPENDIX B: CURRICULUM VITAE

Surname	FRUGIER
First name(s)	Emilie Cecile
Date of birth	August 26th, 1976
Nationality	French, Canadian

Education

Sir Winston Churchill Secondary School (St. Catharines, Canada), 1995, Ontario Secondary School Diploma

McGill University (Montreal, Canada), 2000, Bachelor of Science

Ecole Nationale Supérieure de Chimie de Paris (France), 2002, Diplôme d'Ingénieur Grade de Master, Master's thesis title: "Synthesis of a crude petroleum cracking catalyst based on metal oxide nanoparticles using sol-gel methods"

Eidgenössische Technische Hochschule (ETH) Zürich, 2004, Master of Advanced Studies in Intellectual Property

University of Zurich, 2010-present, Doctoral thesis title: "Computational design and experimental validation of selective inhibitors of enzymes regulating post-translational modifications"

Employment

Omya Development AG (Oftringen, Switzerland), 2004-2010, Intellectual Property Officer
BP Chemicals Research and Technology Center (Lavera, France), 2001, Engineering Intern
Paul Sabatier University of Toulouse III Department of Inorganic Chemistry (France), 1999, Intern
Pulp and Paper Research Center of Canada (Montreal, Canada), 1998, Intern

Grants

Candoc Forschungskredit, 2012, Title: "Finding Protein Targets for Safe Inhibitors"